AN ABSTRACT OF THE THESIS OF

Noa Segall for the degree of Master of Science in Industrial Engineering presented on July 17, 2003.

Title: A Usability Comparison of PDA-Based Quizzes and Paper-and-Pencil Quizzes.

Abstract approved: Redacted for Privacy

J. David Porter

In the last few years, many schools and universities have incorporated personal digital assistants (PDAs) into their teaching curricula, in an attempt to enhance students' learning experience and reduce instructors' workload. One of the most common uses of PDAs in the classroom is as a test administrator. This study compared the usability – effectiveness, efficiency, and satisfaction – of a PDA-based quiz application to that of standard paper-and-pencil quizzes in a university course in order to determine whether it was advisable to invest time and money in PDA-based testing. The effects of computer anxiety, age, gender, and ethnicity on usability were also evaluated, to ascertain that these factors do not discriminate against individuals taking PDA-based tests.

Five quizzes were administered to students participating in an engineering introductory course. Of these, students took two PDA-based quizzes and three paper-and-pencil quizzes. One PDA-based quiz and one paper-and-pencil quiz were compared in terms of their effectiveness, measured as students' quiz scores and through a mental workload questionnaire; their efficiency, which was the time it took students to complete each quiz; and their satisfaction, evaluated using a subjective user satisfaction questionnaire. Computer anxiety was also measured, using an additional questionnaire.

It was hypothesized that the PDA-based quiz would be more effective and efficient than the paper-and-pencil quiz and that students' satisfaction with the PDA-based quiz would be greater. The study showed the PDA-based quiz to be more efficient, that is, students completed it in less time than they needed to complete the paper-and-pencil quiz. No differences in effectiveness and satisfaction were found between the two quiz types.

It was also hypothesized that for PDA-based quizzes, as computer anxiety increased, effectiveness and satisfaction would decrease; for paper-and-pencil quizzes there would be no relationship between computer anxiety and effectiveness and no relationship between computer anxiety and satisfaction. Findings showed an increase in quiz score (increase in effectiveness) and an increase in mental workload (decrease in effectiveness) as computer anxiety increased for both quiz types. No relationship was found between computer anxiety and satisfaction for either paper-and-pencil or PDA-based quizzes.

The final hypothesis suggested that user satisfaction would be positively correlated with effectiveness (quiz score and mental workload) for both PDA-based and paper-and-pencil quizzes. No relationship was found between quiz score and satisfaction for either quiz type. User satisfaction was positively correlated with mental workload, regardless of quiz type.

The usability comparison of paper-and-pencil and PDA-based quizzes found the latter to be equal, if not superior, to the former. The effort students put into taking the quiz was the same, regardless of administration method, and scores were not affected. In addition, different demographic groups performed almost equally well in both quiz types (white students' PDA-based quiz scores were slightly lower than those of the other ethnic groups). Computer anxiety was not affected by the quiz type. For these reasons, as well as other advantages to both students (e.g. real-time scoring) and teachers (e.g. spending less time on grading), PDAs are an attractive test administration option for schools and universities.

A Usability Comparison of PDA-Based Quizzes and Paper-and-Pencil Quizzes

by

Noa Segall

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented July 17, 2003
Commencement June 2004

Master of Science thesis of Noa Segall presented on July 17, 2003.

APPROVED:

Redacted for Privacy

Major Professor, representing Industrial Engineering

Redacted for Privacy

Head of the Department of Industrial and Manufacturing Engineering

Redacted for Privacy

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Noa Segall, Author

ACKNOWLEDGEMENTS

It is difficult to overstate my gratitude to my advisors, Dr. J. David Porter and Dr. Toni L. Doolen. Their continuous guidance, inspiration, and endless support made this work possible. I also wish to acknowledge my committee members for their comments on this thesis. Finally, I would like to thank my husband, Gideon, for his help, patience, and love.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF APPENDICES

A Usability Comparison
of PDA-Based Quizzes and Paper-and-Pencil Quizzes

1. Introduction

The personal digital assistant (PDA) market has grown dramatically in recent years. In 2002 alone, over 12 million handheld devices were sold (Kawamoto, 2003), and a 17.6% annual growth in sales is expected between 2003 and 2006 (Europemedia, 2002). Lower unit prices, improved operating systems, and a wave of multimedia and wireless functionality being integrated into PDAs will contribute to increased PDA sales in years to come (PDA cortex, 2003).

PDAs fall into two major categories: handheld computers and palmtop computers. Handheld computers are generally larger and rely on miniature keyboards and touch screen technology for data entry, while palmtops use touch screens and handwriting recognition programs for input. Most PDAs run one of two operating systems: 3Com's Palm OS or Microsoft's Pocket PC (formerly known as Windows CE).

The first PDAs, launched in the mid-1990s, were Apple's Newton Message Pad and the Palm Pilot. The latter dominated the PDA market for several years. Originally, PDAs were used mostly to manage personal information by offering programs such as calendars and phone books. Today there are numerous PDA brands, and their enhanced capabilities include wireless internet access, games, and the ability to play audio and video files. The PDA market is continuing to develop, with efforts concentrated on making PDAs faster, enlarging their memory, and merging them with cellular phones.

As a result of increasing functionality, PDAs have become pervasive in many work environments, and lately, they have also come into use in educational environments. A study conducted by SRI International, a nonprofit research and development firm, showed PDAs to be useful to students in tasks such as collecting data, writing papers, checking facts, synching data with desktop computers, and collaborating on projects (Dean, 2002). The academic community, encouraged by these findings, is now moving forward and initiating efforts to study the impact of the PDAs on student learning (Dean, 2002).

1.1. Motivation

For the last two decades, the proliferation of computers in education has stimulated the development of many new tools that assist instructors in teaching, evaluating, and directing student learning. Since a common tool to assess student performance is test administration, computer-based testing is one of the more frequently developed applications. Computer-based testing offers several advantages to students, including more accurate grading, immediate performance feedback, real-time scoring, and improved security. Instructors benefit from this type of test administration by spending less time on manual data entry and grading, enabling them to focus on other tasks such as helping weaker students improve their performance. PDAs add relatively low purchase costs to the advantages of using desktop computers in educational settings, and their mobility eliminates the need for a specialized computer lab – they can be used anywhere, including a regular classroom.

As new technologies become available, PDA prices are dropping and the targeted market is shifting from the business community to the general public, including students and teachers. Many schools and universities have incorporated handheld computers into their teaching curricula, in an attempt to enhance students' learning experience. Their uses are varied, from problem solving in class to enabling wireless communication with teachers, friends, and the library anywhere on campus. Frequently, PDAs are also employed in the administration of tests. Instructors who have used handheld computers to administer exams have reported increased enthusiasm on the part of students, although student attitudes have seldom been measured and validated.

Since handheld and wireless technologies are relatively new and continually evolving, there is very little research on the impact of using PDAs to administer tests. Most of the literature is focused on technical development and implementation issues, rather than measuring and analyzing test effects. But until handheld computers and paper-and-pencil are compared as methods for test administration, it remains unknown whether PDA-based tests provide a valid measure of student performance while reducing workload for instructors.

1.2. Objectives

The design of a user interface for any application is of great consequence. A user should be able to achieve predefined goals quickly and easily, and should be satisfied with the product. Usability – making a user interface efficient, effective, and satisfactory – has become an established part of the development lifecycle of many web sites, software programs, operating systems, etc. This is due partly to the benefits of investing in a usable product, but also to the possibly detrimental consequences of overlooking usability.

A usable user interface is crucial when the application is a PDA-based test: the effort students put into taking an exam should be the same, regardless of administration method, and scores should not be affected. If PDA-based tests are shown to be as usable as paper-and-pencil tests, and if factors such as computer anxiety, age, gender, and ethnicity do not discriminate against individuals taking PDA-based tests, then it may be advisable to invest the time and the money to switch to this form of automated testing. Therefore, the objectives of this research are:

- To compare the usability (effectiveness, efficiency, and user satisfaction) of PDA-based tests and standard paper-and-pencil tests.
- To learn if computer anxiety affects the effectiveness of and satisfaction with PDA-based tests or paper-and-pencil tests.
- To assess if a relationship exists between satisfaction and effectiveness for PDA-based tests and for paper-and-pencil tests.
- To determine whether different population groups react differently to PDA-based testing.

1.3. Contribution

The primary objective of this research is to gain an understanding of the impact of using PDAs to administer exams in educational environments. It illustrates the feasibility of administering exams using PDAs and also discusses some potential problems that need to be addressed. The research findings show PDA-based quizzes to be more efficient than paper-and-pencil quizzes and equally effective and satisfactory to students. Moreover, the effects of computer anxiety on effectiveness and satisfaction are the same for both quiz administration methods, and different demographic groups react nearly equally well to both quiz types. Since PDA-based quizzes and paper-and-pencil quizzes are shown to be equivalent, decision-makers in schools and universities should consider the use of PDAs to as an alternative method to administer exams if they aim to lighten instructors' workloads and provide accurate, immediate feedback to students.

2. Literature Review

2.1. PDAs and Human Factors

The design of a graphical user interface (GUI) for PDAs needs to balance two opposing demands:

- The physical demand limits the size of the user interface to that of a small screen.
- The functional limitation requires a sufficiently large interface to show enough information so that the device is actually useful (Kamba, 1996).

Many papers have been published on human-computer interaction with mobile devices, such as PDAs, and a large majority deals with the challenge of the physical and functional demands imposed on interface design. These papers can roughly be categorized into three groups: the first describes methods developed to overcome this problem, the second lists guidelines for small screen GUI designs, and the third discusses specific PDA applications and how the screen size limitation was handled.

2.1.1. Information Visualization

A large part of the research on PDAs in the context of human factors describes different methods for displaying large amounts of information on small screens, mostly for the purpose of web browsing. These tools serve to enhance information visualization, defined as "the use of computer-supported interactive visual representations of abstract data to amplify cognition" (Xerox, 2002). A prominent approach to information visualization, which has been developed into different

applications, is the focus and context visualization technique (e.g. Björk et al., 1999, and McGookin and Brewster, 2001). This technique uses algorithms to divide the information to be presented into two parts: focus and context. Focus is the part of the information that is of greatest interest to the user, therefore it is displayed in full detail. Context is the less relevant information and is displayed in less detail. Björk et al. (1999) used a tile-based representation of the focus and context method to display web pages by applying a technique called flip zooming. In flip zooming, key terms are selected based on their frequency of appearance in the web page. They are then grouped and placed in sequentially ordered discrete displays. One of these displays is in focus, meaning that it is located in the center of the screen and clicking on it will display the regular HTML formatting of the text. The rest of the displays (i.e. the context) surround the focus. Users can select any visible display to become the focus by clicking on it. This concept is depicted in Figure 1(a).

Another method for conserving space on PDA screens entails making control objects, such as icons, semi-transparent. Consequently, they appear to hide beneath the text (see Figure 1(b)). Kamba et al. (1996) experimented with this method by placing text that included hyperlinks above semi-transparent icons, and reported that subjects preferred links to be selected before the icons. That is, if a link was located above an icon, a short stylus click activated the link and a long click activated the icon. A similar tool for presenting information in layers was proposed by Masui et al. (1999). The overlay method visually combines information on two or more separate layers and can be used in computer-aided design (CAD) or to display complex maps (e.g. a topological map and architectural drawings).

Brewster (2002) suggested minimizing buttons and providing auditory feedback when buttons are pressed, in order to allow more room for other applications on the PDA display. He found button size could be reduced from 16×16 to 8×8 pixels

when buttons were sonically enhanced without a great loss of performance. He also looked at subjective workload, "...the effort invested by the human operator into task performance," and found it to increase when button size was reduced.

Taivalsaari (1999) offered another solution to the limited screen size problem: the event horizon user interface model. The model's key principle is that the display can be virtually compressed and expanded by moving objects radially farther away or closer to a sink (the event horizon) in the center of the screen. This concept is illustrated in Figures 1(c) and 1(d). This collapsible interface, similar to scrolling in a desktop computer, allows an unlimited number of objects to be stored, visually organized, and manipulated in a virtually large but physically limited screen.

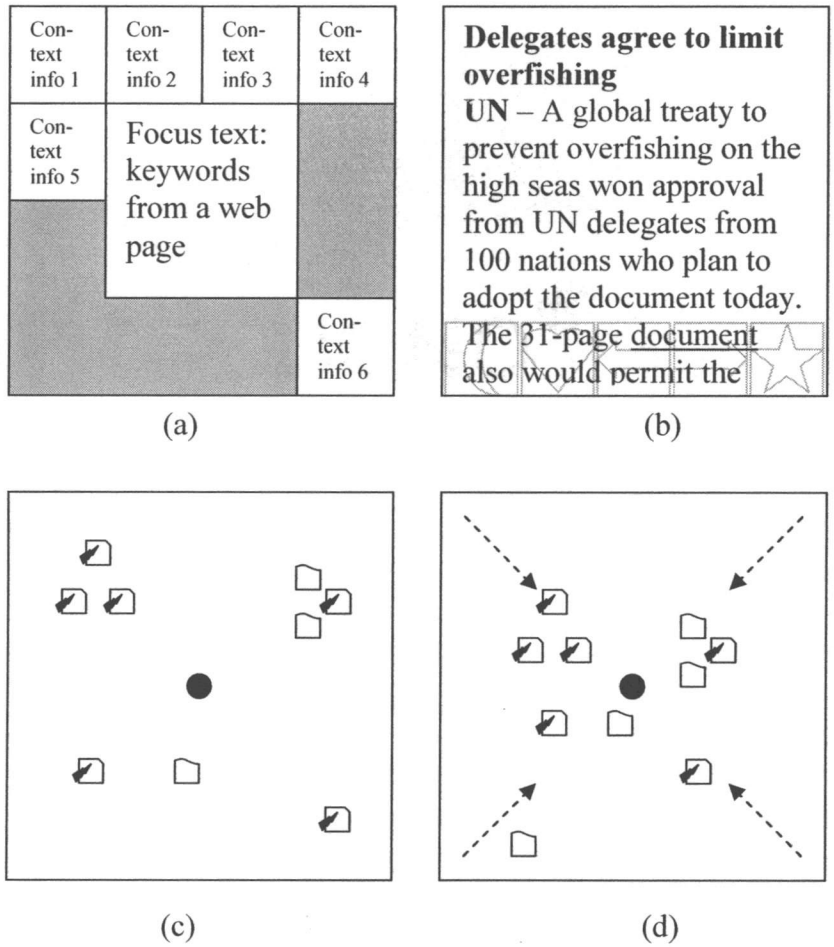


Figure 1: Information visualization techniques, displayed on a PDA screen. (a) Björk et al.'s implementation of focus and context. (b) Semi-transparent control objects used by Kamba et al. (c) and (d): The event horizon, developed by Taivalsaari. (c) Initial layout. The event horizon ("sink") is in the center. (d) Layout after the screen has been "compressed". Objects move closer to the event horizon and may become "sucked in". Objects that were previously outside the screen area may become visible.

Table 1 summarizes the different approaches to information visualization and the techniques used by researchers to evaluate their usability. More advanced methods exist, such as zooming, scaling, and infinitely large virtual displays. However, limited computing power, poor screen resolution, and strict memory constraints make them difficult to utilize (Taivalsaari, 1999). Despite the growing body of

research in this field and the wide range of information visualization methods, no single method, to date, replicates the usability of a desktop computer's user interface.

Table 1: Information visualization methods for PDAs

Method	**Independent Variables**	**Dependent Variables**	**Findings**
A tile-based representation of the focus and context method to display web pages used by applying a technique called flip zooming (Björk et al., 1999)	Presentation method (focus and context browser vs. a regular browser)	User satisfaction	The focus and context browser received higher ratings than a regular browser
Multimodal focus and context, where focus is the visual display and context – a spatial audio display (McGookin and Brewster, 2001)	No experiment was conducted	-	-

Table 1 (continued)

Method	Independent Variables	Dependent Variables	Findings
Control objects, such as icons, are displayed in a semi-transparent format so that they appear to hide beneath text (Kamba et al., 1996)	When a link is located above an icon, (a) whether clicking will select the link or the icon first and (b) the length of the response delay when switching between the layers	Effectiveness (number of errors) and user satisfaction	Subjects preferred links to be selected before the icons; error rates for links-first and icons-first were similar. Subjects preferred a short response delay; error rates for short and long delays were similar
The overlay method: information is presented in overlapping layers (Masui et al., 1999)	Number of layers presented (2-5)	Efficiency (time to complete task) and effectiveness (correct answer rate)	An increase in the number of layers decreased the correct answer rate and increased execution time

Table 1 (continued)

Method	Independent Variables	Dependent Variables	Findings
Minimizing buttons and providing auditory feedback when buttons are pressed (Brewster, 2002)	Button size	Effectiveness and efficiency (number of successful tasks), subjective ratings of workload, annoyance, and user satisfaction	Button size could be reduced from 16×16 to 8×8 pixels when buttons were sonically enhanced without a great loss of performance, but subjective workload increased. Subjects were not annoyed with audible buttons and preferred them to silent ones
The event horizon model: the display can be virtually compressed and expanded (Taivalsaari, 1999)	No experiment was conducted	-	-

2.1.2. Guidelines

Several papers have provided general guidelines for designing user interfaces for handheld mobile devices. Abramovici and Klußmann (1994) present a style guide based on the action theory. According to the action theory, the user forms a conceptual intention, reformulates it into commands, constructs the required syntax, executes the action, and evaluates the outcome (Shneiderman, 1998).

Abramovici and Klußmann (1994) maintain that the user should not be concerned with the means of reaching a given goal, only with what that goal is. User interfaces should therefore be designed in a way that does not assume that the user understands the system's logic. Based on this principle, they developed four guiding principles for user interface style design – guidance, explicit control, adaptability, and homogeneity.

- Guidance is achieved by informing the user – providing feedback on both the system's and the user's actions.
- Guidance should not prevent the user from having explicit control over the system, which means, among other things, that the vocabulary used in the GUI should be familiar to the user.
- Adaptability refers to the user's power to configure the interface as well as to the system's flexibility in taking the user's expertise into account. Novices will require more help, while experienced users will prefer a more direct interface.
- The GUI must be homogeneous – its "look and feel" should be consistent throughout.

Branaghan (2001) talks about three characteristics of successful consumer products: they should be useful, usable, and desirable. For example, for a product to be usable, it should be learnable, efficient, memorable (the interface should not have to be relearned every time it is used), etc. The author maintains that most PDAs are useful and desirable, but not usable. The reason for this is the unsuccessful tradeoff between miniaturization of the physical device and functionality. The industry is attempting to introduce products that pack more and more functions into less and less space. Strategies for managing this tradeoff are suggested. The author proposes that small devices be used only for quick and non-sustained tasks such as sending quick responses to emails, and that activities such as web browsing be left to computers with larger displays. Additionally, functions

should be put together on the same device only if they are used in similar contexts or serve a complimentary function.

Mohageg (1999) compares PDAs to personal computers in terms of the target user and the tasks that each machine needs to fulfill. From this research, three design suggestions are made. The first is to identify the characteristics of PDA tasks, such as the length of the interaction, and to recognize their implications. Since interactions with PDAs are usually short, a sample design implication is that ease of learning is critical. The second suggestion is to adapt PC applications to the PDA – an e-mail application, for example, can be difficult to use when the main input method is a stylus. The final suggestion is to simplify – keep the number of functions as well as the number of choices a user can make down to a bare minimum.

Table 2 provides a summary of the suggested guidelines for PDA user interface design based on previous research. Many of these principles are applicable to the design of generic user interfaces. Some guidelines, such as adaptability and simplicity, are contradictory. It remains to the application developer to balance the opposing requirements.

Table 2: Guidelines for the design of PDA user interfaces

Guideline	**Definition**	**Source**
Guidance	Provide feedback on system and user actions	Abramovici and Klußmann (1994)
Explicit control	Operators should feel that they are in charge of the system and that it responds to their actions	Abramovici and Klußmann (1994)
Adaptability	The system should take the user's experience into account and allow interface configuration	Abramovici and Klußmann (1994)

Homogeneity	Provide a consistent "look and feel"	Abramovici and Klußmann (1994)

Table 2 (continued)

Guideline	Definition	Source
Usefulness	Provide functionality to help the user achieve predefined goals	Branaghan (2001)
Usability	The interface should have the following features: learnability, efficiency, memorability, and error minimization	Branaghan (2001)
Desirability	The interface should evoke a strong emotional reaction from its user	Branaghan (2001)
Identify a target domain	Design interfaces with the task characteristics in mind	Mohageg (1999)
Dedicated devices mean dedicated interfaces	Adapt PC applications to the special needs of the PDA	Mohageg (1999)
Simplicity	Minimize the number of functions and the number of choices the user can make	Mohageg (1999)

2.1.3. PDA Applications

A third factor related to this research is human factors. Human factors examines issues relevant to specific PDA applications. Bellamy et al. (2001) designed an e-grocery application that was deployed on a device with a Palm operating system. Grocery store customers could use this application to do their shopping on the go. In order to evaluate the application's usability, a small number of customers were asked to create and place an order on the PDA while talking out loud and responding to questions. Nyberg et al. (2001) compared three devices in terms of performance – an integrated mobile phone and PDA, a prototype that had both

telecommunications and PDA capabilities, and a PDA alongside a mobile phone. Subjects were asked to carry out several information handling and call handling tasks. Four variables – effectiveness, efficiency, mental workload, and satisfaction – were measured for each of the devices. Effectiveness was evaluated as the number of tasks completed, and efficiency was determined by assessing completion time for each task and the number of keystrokes needed by users to accomplish the task. Mental workload and user satisfaction were measured subjectively using rating scales. The prototype developed by the researchers was generally found to have lower scores on each of the four variables.

Other research has looked at using a PDA alongside an additional device. Rekimoto (1998) utilized a PDA as a tool to address difficulties in interacting with a digital whiteboard. He found that text entry and data handling were cumbersome tasks to accomplish with the whiteboard, and employed a pick-and-drop method to transfer information from a PDA to the digital whiteboard. Robertson et al. (1996) used a PDA to interact with a television that displays real estate information. Users could instruct the television, via the PDA, to display data such as floor plans and pictures of houses. The PDA could also be used as a stand-alone device: on a visit to a selected house, potential buyers could view a map of the neighborhood on the PDA. The usability of these applications was not evaluated.

2.2. PDAs in the Classroom

Since PDAs are relatively new devices on the market, and their wireless communication abilities are even more recent, research on the use of PDAs in the classroom is rather limited. Existing research has concentrated on practical topics having to do with the development and implementation of new educational applications for PDAs. The applications described target a variety of needs – from data collection and online testing to presenting teaching materials. In general, the impact of this technology on student learning has yet to be evaluated.

Cook (2000) describes the National Classroom Project, wherein PDA educational applications were developed for 5th and 9th graders and for college students. He details technology issues he faced when integrating mobile computers into a school environment, such as setting up the devices, ensuring network connectivity, and developing software for special applications. He then describes the design of online tests for students, as well as his in-class experience with implementing them and lessons learned from this experience.

Hudgins (2001), a high school teacher in California, reports her experience as a participant in a pilot program with a handheld device specifically designed for use in classroom testing – the Classroom Wizard from Scantron Corp. These quizzes are not conducted online, rather, a quiz is beamed (i.e. transferred via an infrared link) from a desktop computer to the PDA, after which the student fills it out and beams it back to the computer, which then reports the results to both student and teacher. She notes that time spent on grading is reduced and that students enjoy this testing method. These assertions, however, are not validated or tested for statistical significance.

Shotsberger and Vetter (2001) describe project Numina, a cooperative effort between the University of North Carolina at Wilmington, Pearson Education (Prentice Hall), and Hypercube to integrate wireless mobile technologies into the classroom. One of the applications developed is the Student Response System (SRS, formerly known as SWATT). In this application, the instructor poses a question and directs students to a web site that generates a form on their PDA screens through which they submit their responses. This question and answer format may be used to display student responses as a bar chart via a data projector or to give quizzes. Other applications include using an online version of a textbook and providing chemistry functions to students. In this paper, the authors did not evaluate the effectiveness of the PDA applications which were developed.

Chen, Myers, and Yaron (2000) of Carnegie Mellon University used PDAs to carry out ordinary and concept tests in a chemistry class. A concept test is a test designed to be taken as part of a lecture, with immediate feedback displayed to the students and instructor. Essentially, it is similar to the SRS application described by Shotsberger and Vetter. The authors discuss administrative, hardware, software, and implementation issues, as well as describing the classroom use of the PDAs in detail. At the end of the course, students were asked to respond to a survey regarding their views on the use of different PDA applications, the PDA concept tests (compared to raising hands or flash cards as a response to the lecturer's questions), and PDA characteristics such as screen size. Fifty users, about half of the students in the chemistry class, filled out the questionnaire. Results show that over half of the respondents preferred the PDAs to a show of hands or use of flash cards as a method of conducting concept tests, despite setbacks such as connection problems and batteries that lost power quickly. The authors did not complete any statistical analysis to validate these conclusions.

Kabara et al. (2000) describe current and future usage of mobile devices at the University of Pittsburgh. Currently, students have wireless data access from several parts of the university, enabling them to communicate with friends and instructors and to retrieve information from resources such as electronic databases and laboratory equipment. Future implementations include attending class virtually, taking notes via a pen-based interface, and presenting information to a class without standing at the blackboard. Metrics for evaluating these applications will be both technical (cost considerations and network performance) and educational (the usefulness of these tools for students and instructors).

2.3. Effects of Computer-Based Test Administration

Computer-based testing (CBT) possesses obvious benefits with respect to test administration, such as improved security and accurate, immediate scoring. The primary concerns regarding CBT are whether performance, as measured by test scores, is equivalent to that of paper-and-pencil tests, and whether irrelevant extraneous variables, such as computer anxiety, affect performance. Conclusions from previous research are not consistent. Some studies point to an equivalence of mean achievement scores between test versions, while others show significantly lower results on computer-based tests (Chin and Donn, 1991). In an experiment conducted by Chin and Donn (1991), high school students were given either a computerized or a written version of a science test. CBT scores were actually found to be higher, on average, than paper-and-pencil scores, possibly because students tried harder and were more reluctant to select "I don't know" as an answer with CBT. Other variables that may be affected by test administration method, such as the length of time required to complete each type of test, have not been measured. Student attitudes towards computerized tests, when examined, were often positive, but not statistically verified (Chin and Donn, 1991, Chen et al., 2000, and Hudgins, 2001).

Validity is an important issue when discussing CBT. A widely cited definition of test validity is "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick, 1992). Most research on the validity of CBT relates to psychometric exams such as the Graduate Record Examination (GRE) or professional exams such as the Armed Services Vocational Aptitude Battery (ASVAB). Some validity principles may still be applied to classroom use of this testing method. Validity of computer-based tests, compared with paper-and-pencil tests, is enhanced by

automated scoring, since the dependence on human raters is eliminated. However, construct-irrelevant variance may be a threat to validity in CBT (Huff and Sireci, 2001). For example, computer proficiency may influence test performance if the ability to interact successfully with a computer is necessary to do well on a test that is not designed to measure computer facility. Since social class differences are associated with computer proficiency (Huff and Sireci, 2001), this is an important source of variance to overcome. A lack of familiarity with the computer platform used to administer the test poses a similar problem; these setbacks may also cause test anxiety with certain examinees. Test score validity may also be compromised when the CBT user interface is not designed according to accepted human factors practices (Booth, 1998).

Perkins (1995) looked at the effects of computer-based versus paper-and-pencil tests on computer anxiety and performance. He had students who took his "Computers for Teachers" course take a multiple-choice test on the material and fill out a computer anxiety survey both in the beginning and at the end of the course. A control group took a written version of the tests and surveys, while the experimental group took a written version of the first test, a computerized version of the second test (at the end of the course), and a computerized version of both surveys. In both groups, anxiety decreased and performance (measured as the test scores) increased over the length of the course. No significant difference, however, was found between the groups, thus no effect of computer-based testing on performance and anxiety was established. In addition, a negative relationship was observed between anxiety and performance: lower anxiety predicted higher performance.

In the second part of his research, Perkins also examined the effects of age, gender, and previous computer experience on performance and computer anxiety. Age was found to have no influence on these measures: the mean score of both

undergraduate and graduate students was not significantly different. Females were found to have lower scores than males on the tests, although they showed greater improvement over time. They also exhibited higher anxiety than males, but their anxiety level dropped the most by the end of the study. Students who owned computers or who had previous computer experience outperformed those who did not, and their anxiety level was found to be significantly lower.

Dimock (1991) suggested that differences in performance between computer-based and paper-and-pencil tests could be explained by the different formats used to display test questions. Often, questions in computerized tests are presented one at a time (card format), while in paper-and-pencil tests they are grouped – several questions are visible to the examinee at the same time (booklet format). Dimock performed two experiments. In the first, subjects completed a written version of the Verbal Reasoning part of the Differential Aptitude Tests where questions were presented in either card or booklet format. The booklet format, in which questions were grouped together, was found to be superior to the card format in terms of test scores. In the second experiment, subjects completed the same test with questions presented in the card format. The test was administered either on a computer or on paper. The paper-and-pencil version of the test was found to be superior; computer anxiety and computer familiarity, as measured by questionnaires, could not explain the differences in test performance.

Table 3 presents a summary of research findings on the effects of test administration method on performance. Of the research examined, results are mixed. Chin and Donn (1991) found students' scores to be higher on CBT, Dimock (1991) found them to be lower, and Perkins (1995) found no difference between scores on computer-based and paper-and-pencil tests. These findings are in accordance with Chin and Donn's (1991) literature review. The existence of a relationship between test administration method and performance is not certain.

Table 3: Summary of research findings on CBT effects

Author(s)	Measured Factors	Effect on Performance
Chin and Donn (1991)	Test scores – CBT vs. paper-and-pencil tests	Higher scores on CBT
	Gender	Test administration method did not affect performance
	Test anxiety	Test administration method did not affect performance
Perkins (1995)	Test scores – CBT vs. paper-and-pencil tests	No significant difference
	Computer anxiety	Test administration method did not affect performance
	Gender	Males received higher test scores on CBT
	Age	Test administration method did not affect performance
	Computer experience	Experienced users received higher test scores on CBT
	Computer ownership	Computerowners received higher test scores on CBT
Dimock (1991)	Test scores – CBT vs. paper-and-pencil tests (questions presented one at a time)	Lower scores on CBT
	Computer anxiety	Test administration method did not affect performance
	Computer experience	Test administration method did not affect performance

2.4. Computer Anxiety

Computer anxiety is defined as "the complex emotional reactions that are evoked in individuals who interpret computers as personally threatening" (Raub, 1981). Igbaria et al. (1994) describe these emotional reactions as phobias, uneasiness, or apprehension. Computer anxiety has been examined frequently in research on computers in education in an attempt to understand whether it may discriminate against certain user groups. Studies regarding the existence of a relationship between computer anxiety and performance in tests have had mixed results. Some have shown high anxiety to predict low test scores in CBT, while others showed no difference in computer anxiety and test scores between computer-based and paper-and-pencil based testing. Chin and Donn (1991) suggested that if the tasks to be performed during the test are kept simple, computer anxiety would not be a significant factor.

Many empirical studies have attempted to identify demographic factors that correlate with the occurrence of computer anxiety. However, findings across this body of literature seem to be contradictory. Regarding gender, some studies indicate that females experience more computer anxiety than males, others show that males are subject to greater levels of anxiety than females, and a third group found no significant difference between males and females in the extent to which they experience computer anxiety (Worthington and Zhao, 1999). Maurer (1994), in a literature review of computer anxiety research, points out that findings showing females to experience more computer anxiety than males are problematic, since research has found males to have greater access to computers and more computer experience.

The results of many studies exploring age effects demonstrate similar inconsistencies. For example, Gilroy and Desai (1986) found no correlation between age and computer anxiety, while Bowers and Bowers (1996) observed a positive relationship in one college social science class and no relationship in another. Maurer (1994) also found some evidence to show no effect due to age and some to show that younger subjects tended to be less anxious.

The influence of ethnicity on computer anxiety has not been investigated thoroughly. Bowers and Bowers (1996) did not find any correlation between race and computer anxiety in undergraduate social science students. Gilroy and Desai (1986) obtained similar results for a population of undergraduate and graduate students.

Igbaria et al. (1994) looked at a population of managers and professionals to determine the effects of computer anxiety on perceived usefulness, perceived fun, satisfaction, and system usage in the workplace. Perceived fun was defined as the system being pleasant, enjoyable, interesting, etc., while satisfaction was measured by asking participants to rate the system's quality of display, speed of response, etc. They found perceived usefulness (a subjective evaluation of effectiveness) and perceived fun to have a negative relationship with computer anxiety. Computer anxiety was also found to have a negative relationship with both satisfaction and usage, manifested both indirectly (through perceived fun and usefulness) and directly.

2.5. Usability

2.5.1. Measuring Usability

Usability is defined by IEEE (1990) as "the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component." There are many tools for measuring the usability of a software program, web site, or any application that has a user interface. A list of several common methods follows.

- In heuristic evaluation, several (preferably 3-5) human-computer interaction (HCI) experts inspect the interface, using several possible scenarios, against a list of commonly accepted principles. There are ten commonly used heuristics, such as consistency and error prevention.
- Questionnaires and surveys provide structured answers to questions formulated by researchers; interviews and focus groups may be used when unstructured answers are required. Users may be queried on their experience with using a finished product or at the initial stages of development, to extract product requirements.
- Cognitive walkthrough is a review technique where expert evaluators construct task scenarios from an early prototype and then role-play the part of a user working with the interface. Each step the user would take is scrutinized; for example, impasses where the interface blocks the user from completing the task indicate that the interface is missing something.
- Standards inspections ensure compliance with industry standards. One such standard is ISO 9241 (Ergonomic requirements for office work with visual display terminals), DIS Part 10 (Dialogue Principles). This part of the standard specifies a set of dialogue design principles for form-based entries as well as command languages and direct manipulation. Its objective is to optimize dialogue design in terms of effectiveness, efficiency, and satisfaction. An

example of a design principle listed here is controllability, which is attained by allowing the user to determine the interaction speed with the software, enabling a last step undo, etc. (Smith, 1996). In most cases, the inspection is performed by an expert in the standard.

2.5.2. Usability Metrics

A product's usability is often measured by looking at three metrics: effectiveness, efficiency, and user satisfaction. Effectiveness is defined as how well the user achieves the goals he or she set out to achieve using the product. A common measure of effectiveness is the number of errors made by the user while attempting to accomplish a task. Another way to assess effectiveness is to quantify the physical or mental effort put into the task. For HCI tasks, this effort is known as mental workload. Sanders and McCormick (1993) define the idea of mental workload as "a measurable quantity of information processing demands placed on an individual by a task." This concept builds on resource models that postulate a limited quantity of resources available to perform a task. One of the objectives of measuring workload is to compare alternative task designs in terms of the workloads imposed.

Efficiency is the resources consumed in order to achieve a goal. Time is the resource of greatest interest to HCI experts:an efficient task will consume less of the user's time. Therefore, efficiency measurements include time to complete a task, time to learn how to perform a task, time spent on recovering from errors, etc.

Satisfaction is how the user feels about the use of the product. This is a subjective measure, evaluated through user feedback in the form of questionnaires, surveys, etc. There is theoretical and empirical support of the presence of a causal relationship between usage and satisfaction: satisfaction stimulates usage (Igbaria et al., 1994). Consequently, user acceptance of a system may influence its success

or failure. A system's perceived fun was found by Igbaria et al. (1994) to have a positive relationship with perceived usefulness (a subjective evaluation of effectiveness).

Following is a list of potential usability metrics used to evaluate effectiveness, efficiency, and satisfaction.

Table 4: Usability metrics (taken from Mayhew, 1992, Nielsen, 1993, and Whiteside et al., 1988)

Metric	**Evaluation Method**
Effectiveness	Number of errors
	Percent of tasks completed
	Ratio of successes to failures
	Workload (mental or physical)
	Number of features or commands used
Efficiency	Time to complete a task
	Time to learn
	Time spent recovering from errors
	Number of errors
	Frequency of help or documentation use
	Number of repetitions or failed commands
User satisfaction	Rating scale for usefulness of the product or service
	Rating scale for satisfaction with functions and features
	Number of times user expresses frustration or anger
	Rating scale for user's perceived control

3. Research Objectives

Usability is defined as "the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component" (IEEE, 1990). In this research, PDA-based and paper-and-pencil quizzes were evaluated and compared for usability, as measured by effectiveness, efficiency, and satisfaction. In addition, computer anxiety, which previous research has shown to influence performance, was assessed. The variables that were evaluated and their method of evaluation are as follows:

Table 5: Variables

Metric	Evaluation Method
Effectiveness	Quiz score
	Mental workload questionnaire
Efficiency	Time to complete quiz
Satisfaction	Questionnaire
Computer anxiety	Questionnaire

The influence of demographic factors – age, gender, and ethnicity – on the dependent variables was also evaluated and analyzed to identify confounding relationships.

3.1. Research Hypotheses

3.1.1. Hypothesis 1

The method of quiz administration (PDA-based versus paper-and-pencil) directly affects all usability factors: effectiveness, efficiency, and satisfaction. Effectiveness, measured as the students' quiz scores and mental workload, is higher for PDA-based quizzes than for paper-and-pencil quizzes. PDA-based quizzes are also more efficient than paper-and-pencil quizzes as measured by quiz completion time. Finally, user satisfaction, measured using a survey, is higher for the PDA-based quiz than for the paper-and-pencil quiz.

3.1.2. Hypothesis 2

Both effectiveness and user satisfaction with PDA-based quizzes are negatively correlated with computer anxiety: as computer anxiety increases, effectiveness and satisfaction decrease. No relationship exists between computer anxiety and effectiveness in paper-and-pencil quizzes. Likewise, no relationship exists between computer anxiety and user satisfaction in paper-and-pencil quizzes.

3.1.3. Hypothesis 3

User satisfaction is positively correlated with effectiveness for both PDA based and paper-and-pencil quizzes.

3.2. Proposed Path Diagram

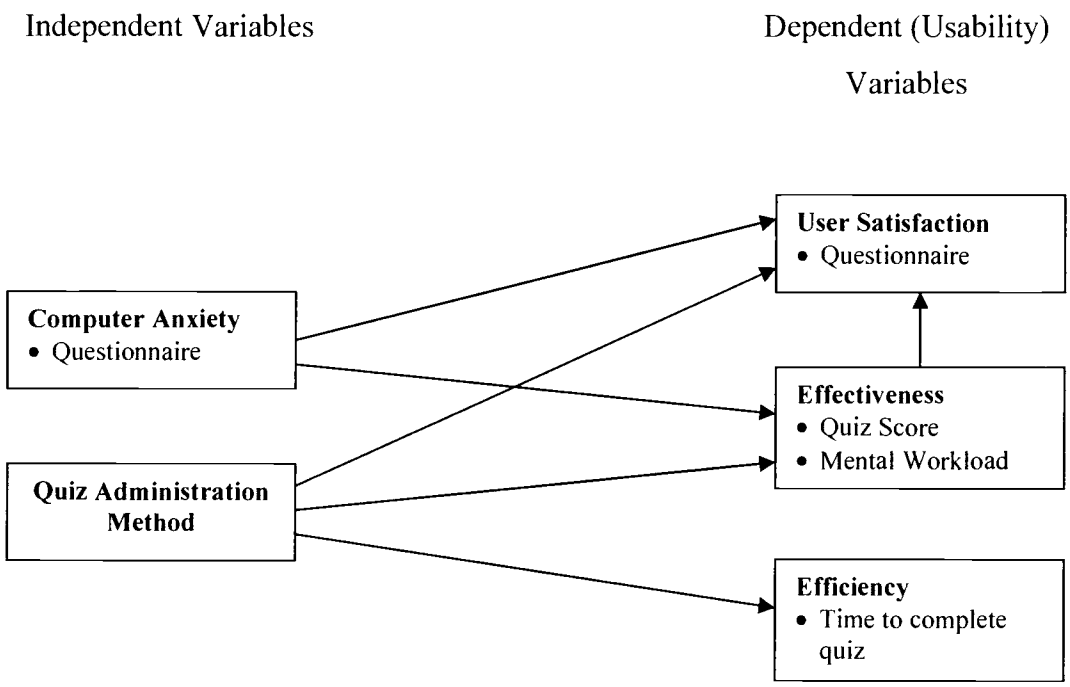


Figure 2: Proposed path diagram

4. Research Methodology

4.1. Dependent Variables

The dependent variables in this study are the three fundamental usability metrics: effectiveness, efficiency, and user satisfaction. These were measured as described in Table 6.

Table 6: Dependent variable measures

Metric	Evaluation Method
Effectiveness	• Quiz score • Five of the six NASA Task Load Index (TLX) scales
Efficiency	• Time to complete quiz
Satisfaction	• Subsections from the Questionnaire for User Interface Satisfaction (QUIS)

4.1.1. Effectiveness

Effectiveness is defined as how well the user achieves the goals he or she set out to achieve using a product. For a PDA-based quiz to be effective, the effort students put into it and the outcome – their grades – should not differ from the effort and grades measured for a paper-and-pencil quiz. Accordingly, effectiveness was assessed by looking at students' efforts and quantified using a mental workload scale and quiz scores.

4.1.1.1. Quiz Scores

When the effectiveness of a user interface needs to be evaluated, researchers often measure users' success or failure at task completion. For example, Masui et al. (1999) looked at the correct answer rate to questions and Brewster (2002)

measured the number of successful tasks subjects were able to finish. Kamba et al. (1996), on the other hand, measured errors made by subjects. In this study, the goal to be achieved is successful completion of a quiz, thus the quiz score was used to quantify effectiveness. This approach is supported by studies comparing computer-based tests to paper-and-pencil tests, where one of the most important variables measured by researchers was test scores (Chin and Donn, 1991, Dimock, 1991, and Perkins, 1995).

4.1.1.2. Mental Workload

Mental workload is often used to assess effectiveness (e.g. Brewster, 2002). Reid and Nygren (1988) state that "...the essence of the major [workload] theories is that the human information processing system has a finite capacity or capacities, and different task situations require varying degrees of capacity expenditure." It is very difficult, if not impossible, to quantify capacity expenditure during task performance. However, several techniques have been developed that attempt to measure workload for a variety of situations.

There are several acceptable measures of mental workload, such as physiological measures, e.g. brainwave activity, and subjective measures, which require users to complete a valid workload scale (Sanders and McCormick, 1993). Although subjective measures may appear inferior since they are not impartial, they have many advantages: they are more direct, less intrusive (since the user is not disturbed during task performance), inexpensive, quick to administer, and finally, they do not require special equipment (Hill et al., 1992). Three workload scales have been researched and validated extensively: The NASA Task Load Index (TLX), the Subjective Workload Assessment Technique (SWAT), and the Modified Cooper-Harper (MCH). Hill et al. (1992) compared these scales, as well as the less known Overall Workload (OW) scale, along four dimensions:

- Sensitivity (factorial validity) – how well each scale is able to discriminate among different workload levels;
- Operator acceptance – users' reactions to the scales;
- Resource requirements – time to complete the scale and time required for training, preparation, and data reduction; and
- Special procedures – time required to customize the scales.

The NASA-TLX was found to be the most sensitive, followed by the OW and SWAT. The NASA-TLX was also the most liked and best in its ability to represent workload (as determined by the operators). In terms of resource requirements, the NASA-TLX took the longest to complete (mean of 51.3 seconds) and, along with the SWAT, demanded more time for data reduction and analysis. These two scales also required subjects to perform a special sorting procedure which, in the case of SWAT, caused many of them to err in their first attempt to perform it. The OW took the least time to complete and required no special resources or procedures. In turn, the NASA-TLX and the SWAT provide additional information that may be used diagnostically to locate and alleviate excessive workload.

Since the NASA-TLX has been widely used and shown to be superior in terms of sensitivity and operator acceptance (Hill et al., 1992) as well as reliability (NATO, 2001), it was used in this study to measure mental workload. This scale is based on the assumption that workload is a hypothetical construct that represents the cost incurred by a human operator to achieve a certain level of performance (NATO, 2001). The NASA-TLX consists of six subscales: mental, physical, and time demands, performance, effort, and frustration (see Appendix A). For the purpose of this study, the scale used to evaluate physical demand was removed (since no physical effort is required to take a quiz), leaving five subscales. Each of these subscales consists of a hundred-point scale divided into twenty, 5-point interval steps. The endpoints have verbal descriptors. Another important aspect of the

NASA-TLX is the development of an individual weighting procedure for combining the results of the different subscales to reduce between-subject variability. Several studies have shown the weighting procedure to be ineffective and recommend ignoring it (NATO, 2001, and Hill et al., 1992).

As stated before, the NASA-TLX has been extensively researched in terms of validity and reliability. Hill et al. (1992) measured its factorial validity, which describes how well the variable that the survey purports to measure can be used to summarize relationships between item responses. They performed factor analysis on data collected from five experiments in which the NASA-TLX was taken by operators of military systems. They found a single factor solution, supporting the view that the scale quantifies a single common factor. Factor loadings varied between 0.899 and 0.942, signifying high factorial validity.

Battiste and Bortolussi (1988) assessed the NASA-TLX intraclass coefficient, a measure of test-retest reliability. This type of reliability is quantified by giving the same survey to a single group of subjects twice. The underlying rationale is that if the survey reflects some meaningful construct, it should assess that construct comparably on both occasions (DeVellis, 1991). Battiste and Bortolussi (1988) found the NASA-TLX to be highly reliable, with an intraclass coefficient of 0.769.

4.1.2. Efficiency

Efficiency is defined as the resources consumed in order to perform a task. In this study, the task is the completion of a quiz, and one of the resources required to achieve this task is time. Thus, efficiency was measured as the time required by the students to take the quiz. Other studies have also used time to complete a task as a measure of efficiency (e.g. Masui et al., 1999), although task success / failure has also been used to this end (Brewster, 2002). The time taken to complete the

paper-and-pencil quiz was self-reported by the students, whereas the time taken to complete the PDA-based quiz was collected electronically.

4.1.3. Satisfaction

User acceptance of software applications is often appraised subjectively using a questionnaire. This is due to the fact that surveys are inexpensive and easy to apply (Root and Draper, 1983). If a questionnaire finds an application meets user needs in an easy and efficient manner, this may be an indication of the application's success; if it finds fault with the application, the questionnaire will provide designers with information that can be used to improve it (LaLomia and Sidowski, 1990).

One commonly used questionnaire for which validity and reliability have been established is the Questionnaire for User Interface Satisfaction (QUIS, Chin et al., 1988). QUIS was created to gauge the user's perception of software usability as it is expressed in specific aspects of the interface (Harper et al., 1997). Version 5.5 of the QUIS consists of 27 items, each of which has a rating scale from 1 to 9 anchored at endpoints with adjectives (e.g. difficult/easy). Items pertain to overall user reactions to the system, screen characteristics, terminology and system information, learning, and system capabilities (see Appendix A).

Internal consistency reliability is the extent to which a survey's items are homogeneous, that is, they all measure the same variable. A widely used measure of internal consistency reliability is Cronbach's alpha coefficient (SPSS Inc., 1999). Cronbach's alpha is a value from 0 to 1 defined as

$$\alpha = \frac{k \cdot \overline{cov} / \overline{var}}{1 + (k-1) \cdot \overline{cov} / \overline{var}}$$

where k is the number of items in the survey, $\overline{cov}$ is the average inter-item covariance, and $\overline{var}$ is the average item variance. Alphas above 0.7 are usually

considered acceptable proof of a survey's reliability (Nunnally, 1978). For QUIS, the alpha coefficient was found to be 0.939 (Chin et al.. 1988).

A survey that has high construct validity is one that measures the variable it was intended to measure. To evaluate QUIS's construct validity, Harper et al. (1997) correlated item scores with the six general satisfaction items (items 1-6 in Appendix A) validated in previous studies. They found mean correlations between each main item and the general satisfaction scale to range between 0.49 and 0.61, suggesting a good agreement between the different parts of QUIS and general satisfaction while not being redundant.

Several modifications were made to the QUIS in order to incorporate it into this study (see Appendices B and C). First, the number of questions was reduced to only those most relevant to the quiz application. This was done to shorten the length of the survey. Thus, one of the six general satisfaction items, which pertained to the application's power, was removed. Items from the other parts of the survey were also removed, reducing it from 27 to 14 items in the version evaluating the PDA-based quiz. For the version used to evaluate the paper-and-pencil quiz, only the first five general satisfaction items were used. Second, survey items were changed to a question format. For example, the screen item "Sequence of screens" used answers that varied from "confusing" to "very clear". This was changed to "Was the screens' sequence confusing or clear?" Finally, the rating scales were reduced from nine to five points, since studies have shown that there is little prediction and statistical power to be gained by using more than five points. In addition, respondents find it hard to distinguish between more than five scale categories (Devlin and Dong, 1993).

4.2. Independent Variables

The independent variables in this study were the quiz administration method (PDA versus paper-and-pencil) and computer anxiety. In addition, demographic information was collected from the research participants for a post-hoc analysis of the data. The independent variables were measured as described below.

4.2.1. Quiz Administration Method

The effect of the quiz administration method on usability was evaluated. Two methods were tested: paper-and-pencil and PDA-based quizzes.

4.2.2. Computer Anxiety

One previously developed instrument used for measuring computer anxiety is the computer anxiety subscale of the Computer Attitude Scale (CAS) developed by Loyd and Gressard (1984). Cronbach's alpha for this subscale was reported to be 0.86 in Loyd and Gressard's (1984) study of eighth through twelfth-grade students and 0.89 in Gressard and Loyd's (1986) study of school teachers. The ten statements, taken from Gressard and Loyd (1986, with permission from D. Loyd, personal communication, February 2003), are included in Appendix A. Research participants use a four-point Likert scale to report the extent to which they agree or disagree with the statements. For the purpose of this study, the number of items was reduced to five and the number of points in the Likert scales was increased from four to five, with 1 signifying "strongly disagree", 3 signifying "neutral", and 5 signifying "strongly agree" (see Appendices B and C).

Loyd and Gressard (1984) and Gressard and Loyd (1986) examined the factorial validity of the CAS and its subscales. Gressard and Loyd (1986) found a three-factor solution that accounted for 54% of the total variation in the three subscales

that make up the CAS. The five items selected from the computer anxiety subscale for this study were all found to load on the same factor, confirming that they all measure a single construct. Their factor loadings had values above 0.5. Kline (1994) states that factor loadings are considered high, indicating high factorial validity, if they are greater than 0.6 and moderately high if they are above 0.3. Thus, the five items have moderately high factorial validity. These results are consistent with the findings of Loyd and Gressard (1984).

4.2.3. Demographic Variables

Participants were asked to report their age group (whether they were older or younger than 23), gender, and ethnicity. These variables were measured since they could presumably have an influence on the different aspects of usability measured in this study. In addition, although several papers on computer-based testing have looked at the effects of different demographic variables on effectiveness (i.e. performance, as measured by test scores; e.g. Chin and Donn, 1991, and Perkins, 1995), the topic of PDA-based testing has received little attention. The literature on this subject has mostly examined technical issues surrounding the implementation of such tests (e.g. Cook, 2000), and when age or gender are discussed, no statistical analysis of their effects take place (e.g. Chen, Myers, and Yaron, 2000).

4.3. Survey Reliability Evaluation

Two surveys were designed with items taken from existing questionnaires that measure mental workload (NASA-TLX), computer anxiety (CAS), and user satisfaction (QUIS). One survey was designed to measure students' reactions to paper-and-pencil quizzes and the other was designed to measure student reactions to PDA-based quizzes. The surveys contained only part of the items included in the original questionnaires from which they were taken and the wording of some of these items was altered to fit this study. Therefore, it was necessary to evaluate the internal consistency reliability (the extent to which all survey items measure the same variable) of the surveys in a pilot study. This was done by administering the surveys to a group of students taking a course that used both paper-and-pencil and computer-based quizzes throughout a ten-week academic term. Cronbach's alpha was used as the reliability metric (see section 4.1.3.).

4.3.1. Pilot Study Participants

The participants for this study were College of Engineering students or students considering engineering as a possible degree option at Oregon State University. These students were taught engineering problem solving using computers, how to use spreadsheet tools (Microsoft Excel), and basic programming skills (Microsoft Visual Basic) in Engineering Orientation II, an introductory course for freshmen. The number of participants was 65 in the survey following the computer-based quiz and 70 in the survey following the paper-and-pencil quiz, out of a class of 76 students. In this class, 99% of the students were under 23 years of age, 83% were males and 77% were Caucasian. Prior to taking the survey, an informed consent form was distributed and read to the participants; this form stated that participation in the survey was voluntary and that the survey was anonymous. Students were not compensated in any way for their participation in the survey.

4.3.2. Instrument

The survey for the computer-based quizzes, which consisted of the same scales as those used in the full study, measured subjective user satisfaction, mental workload, and computer anxiety (see Appendix B). Participants were also asked to specify their gender, age group, and ethnic identity. User satisfaction was measured using 14 items taken from the QUIS. Of these, seven measured overall satisfaction (items 1-5, 11, and 14). Overall satisfaction questions included:

- Was the experience of taking the quiz on a computer terrible or wonderful?
- Was it difficult or easy to take the quiz on a computer?

The seven remaining user satisfaction items focused on system characteristics (items 6-10, 12, and 13). System characteristics questions included:

- Was the organization of the screen design confusing or clear?
- How difficult or easy was it to correct your mistakes?

The 14 user satisfaction items were measured on a scale of 1 to 5, with 1 (on the left) indicating low satisfaction and 5 indicating high satisfaction. For the question "How difficult or easy was it to correct your mistakes?" for instance, the scale extended from "Difficult" (1) to "Easy" (5).

Mental workload was quantified using five of the six NASA-TLX items, with the item evaluating physical demand removed (items 15-19). For example, the mental demand item was phrased as follows:

- Overall, how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

The mental workload items were measured on an unnumbered hundred-point scale divided into twenty steps, with the left indicating a low workload and the right indicating a high workload. Thus, the mental demand scale extended from "Low Mental Demand" on the left, to "High Mental Demand" on the right.

Computer anxiety was measured using five items taken from CAS (items 20-24). Computer anxiety statements included:

- I feel comfortable working with a computer
- Working with a computer makes me very nervous

The degree to which participants agreed or disagreed with each statement was quantified using a five-point Likert scale in which 1 (on the left) signified "Strongly Disagree", 2 signified "Somewhat Disagree", 3 signified "Neutral", 4 signified "Somewhat Agree", and 5 signified "Strongly Agree". Strongly agreeing with statements such as "Working with a computer makes me very nervous" would indicate a high level of computer anxiety, while strongly agreeing with "I feel comfortable working with a computer" would indicate low computer anxiety.

The survey for the paper-and-pencil quizzes was comprised of a total of ten items. Five items (1-5) evaluated overall user satisfaction (items 11 and 14 from the computer-based quiz survey were not used) and the other five (6-10) quantified mental workload. Participants were again asked to specify their gender, age, and ethnicity (see Appendix B).

4.3.3. Procedure

Five quizzes were administered at the beginning of the class session in weeks 2, 4, 6, 7, and 9 of the ten-week term. Students took two computer-based quizzes and three paper-and-pencil quizzes. After the completion of their second computer-based quiz (in week 7) and third paper-and-pencil quiz (in week 9), students were asked to fill out the corresponding survey.

4.3.4. Results

Cronbach's alpha for the survey following the computer-based quiz was 0.88 for all 14 items measuring user satisfaction. Alpha was 0.83 for the seven items evaluating satisfaction with system characteristics and 0.77 for the seven items evaluating overall satisfaction. When only the five items corresponding with those measuring overall satisfaction in the paper-and-pencil quiz survey were examined, alpha was found to be 0.75. For mental workload and computer anxiety, Cronbach's alpha was determined to be 0.77 and 0.75, respectively. These results are summarized in Table 7.

In the survey for paper-and-pencil quizzes, Cronbach's alpha for the five items measuring overall satisfaction (out of the seven items used in the survey following the computer-based quiz) was 0.84. For mental workload, using the same items as those used in the survey for the computer-based quiz, the coefficient alpha reliability was found to be 0.88. These results are summarized in Table 7.

Table 7: Pilot survey results: Cronbach's alpha

	Computer-Based Quiz*	**Paper-and-Pencil Quiz****
User satisfaction (14 items)	0.88	NA
Overall satisfaction (7 of the 14 items)	0.77	NA
Overall satisfaction (5 of the 7 items)	0.75	0.84
Satisfaction with system characteristics (7 of the 14 items)	0.83	NA
Mental workload (5 items)	0.77	0.88
Computer anxiety (5 items)	0.75	NA

* n = 65
** n = 70

Table 8 presents the mean scores and standard deviations for the surveys following both the paper-and-pencil quiz and the computer-based quiz. A Mann-Whitney two-sample test was carried out to compare the data from the two surveys where a comparison could be made. The Mann-Whitney is the non-parametric complement to the two-sample t-test. A non-parametric statistical test was used here since prior research has shown that many attitude metrics violate the normality assumption (Besterfield-Sacre et al., 1999) on which the t-test relies. Non-parametric tests do not assume that the data is normally distributed. They are more conservative than tests that assume normality, therefore if they yield a significant difference, a normality-based test will also likely be significant (Besterfield-Sacre et al., 1999).

The overall satisfaction, as measured by five items, was lower for the paper-and-pencil quiz than for the computer-based quiz ($Z = -11.359$, one-sided p-value < 0.001). In addition, the mental workload was higher for the paper-and-pencil quiz than for the computer-based quiz ($Z = -9.598$, one-sided p-value < 0.001). These results are summarized in Table 8.

Table 8: Pilot survey results: Mean scores and standard deviations

		Computer-Based Quiz*		Paper-and-Pencil Quiz**		
	Scale	**Mean**	**STD**	**Mean**	**STD**	**p-value**
User satisfaction (14 items)	1-5	4.24	0.96			
Overall satisfaction (7 of the 14 items)	1-5	4.11	1.07			
Overall satisfaction (5 of the 7 items)	1-5	3.96	1.06	3.01	1.08	<0.001
Satisfaction with system characteristics (7 of the 14 items)	1-5	4.37	0.82			
Mental workload (5 items)	0-100	30.05	23.17	50.05	26.55	<0.001
Computer anxiety (5 items)	1-5	4.28	0.97			

* n = 65
** n = 70

4.3.5. Conclusions

All reliability alpha coefficients for both surveys were above 0.7, the minimum recommended by Nunnally (1978). Thus, these surveys were found to be adequate for the study comparing PDA-based and paper-and-pencil quizzes. Additionally, the computer-based quiz was found to be superior to the paper-and-pencil quiz in terms of overall satisfaction and mental workload.

4.4. PDA-Based Quizzes

Two quizzes were developed for a freshman-level engineering course that run on a handheld personal computer (model HP Jornada 720). This device is depicted in Figure 3. The operating system that runs on the handheld PCs is Microsoft Windows CE. The quizzes were written in Microsoft eMbedded Visual Basic 3.0, a programming language used to develop applications for Windows CE-based devices.

The first and second quizzes consisted of four and nine questions, respectively. The quizzes were saved as .cab files, where .cab is a Microsoft file format that facilitates the efficient compression of multiple files into a single cabinet file (Borland Developer Network, 1999). This is the only file format that enables applications to be downloaded off a web page and installed on a PDA.

Figure 3: HP Jornada 720 (taken from www.hp.com, 2002)

When it was time to take a quiz, students were directed to a location in the class website where they could download the quiz application and install it on their handheld PCs. When students opened the quiz application, they were prompted for

their first and last names and the last four digits of their social security number (see Figure 4).

Figure 4: Login screen

After the students clicked on the OK button, the application first checked that the quiz had not already been taken. When the quiz application was run, an empty text file was created and placed in the PDA (without the student's knowledge) before it terminated. If the student attempted to retake the quiz, the application would run again and check whether this file existed. Its presence would indicate that the quiz had already been taken. The application would inform the student of this (Figure 5) and then terminate.

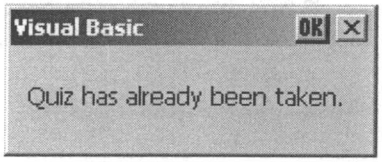


Figure 5: Message box displayed when student attempts to retake the quiz

If the student had not previously taken the quiz, the application would display the first screen of quiz questions. Each individual question and associated responses were presented on a separate screen, as depicted in Figure 6. Each screen also displayed the question's relative position within the quiz (e.g. question 1 of 9) and

the number of points assigned to it. Students were able to move forward and back between the questions. In each question, they were asked to select between several possible answers using option buttons, check boxes, or combo boxes (see Figures 6 and 7).

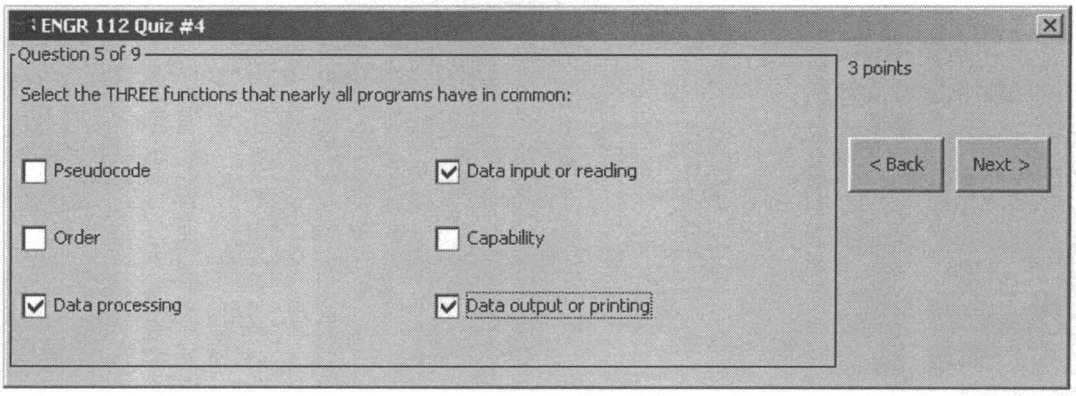


Figure 6: Sample question using check boxes

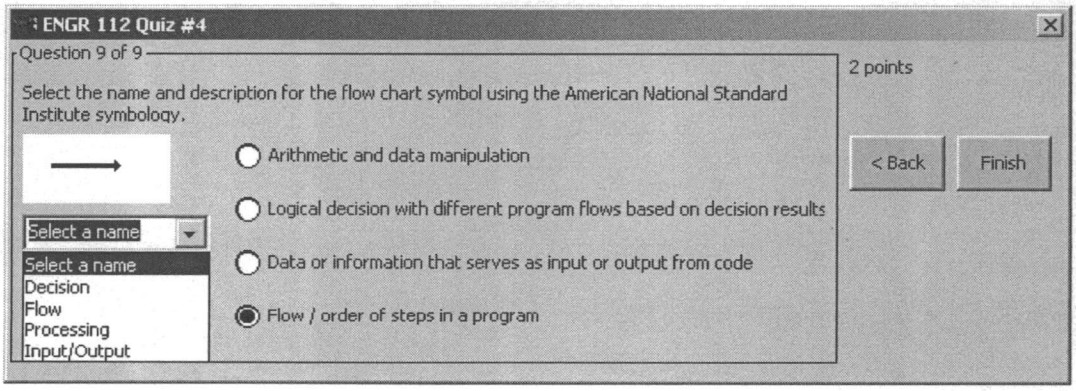


Figure 7: Sample question using a combo box (left) and option buttons (right)

When students reached the final question, the caption on the right-most button changed from Next (Figure 6) to Finish (Figure 7). If the Finish button were clicked, the application would go through each question, beginning with the first, and confirm that it was answered. It would notify the student of the first question it found that was not answered (Figure 8) and return to that question. Consequently, students could not finish the quiz without answering all the questions. If the Finish

button were pressed when all questions were answered, a message box would appear, asking the students to confirm that they did not wish to return to the quiz (Figure 9).

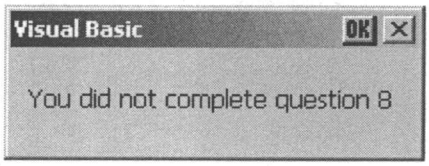


Figure 8: Message box informing the student of not having answered a question

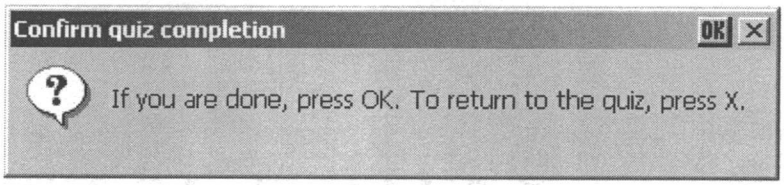


Figure 9: Message box confirming that the student has completed the quiz

When the students confirmed that they were done, their score was calculated and displayed (Figure 10(a)). Each student's score, as well as name, social security number, answers to each question, and the times at which the quiz was begun and finished, were written to a comma delimited text file (.csv format) which he or she was then asked to submit via the web. This file format can be opened by the desktop version of Microsoft Excel, but not by the Windows CE Excel. As a result, students could not view or change this file. When the first of the two PDA-based quizzes was administered in class, seven of the 36 students that took the quiz submitted the wrong file, apparently because other files with similar names were mistaken for the correct file (quiz1.csv). Consequently, the file was renamed using a different format (first name_last name.csv, e.g. John_Doe.csv) for the second quiz (Figure 10(b)).

Students were given the option of seeing the correct answers to the quiz (see Figure 11). The quiz application is further described using a flow chart in Figure 12.

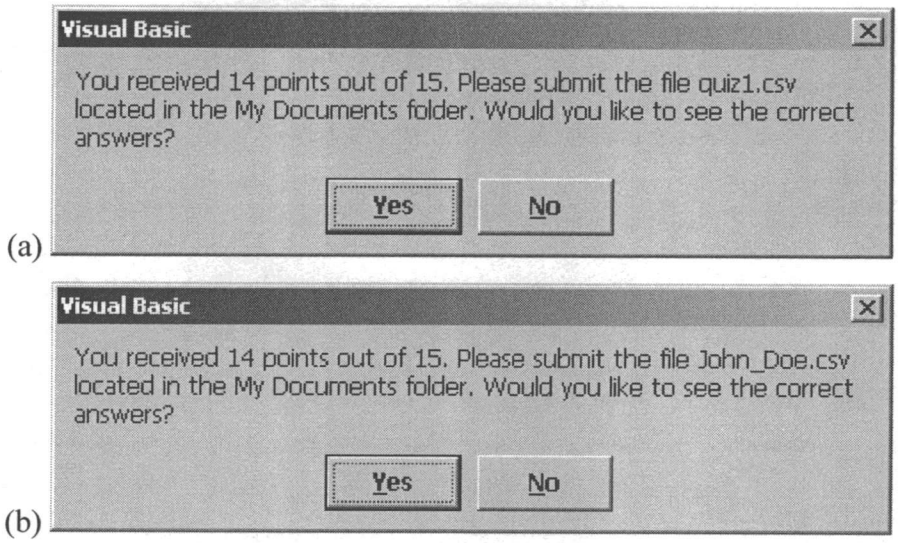


Figure 10: Quiz score. (a) First quiz: file name is quiz1.csv. (b) Second quiz: file name is first name_last name.csv.

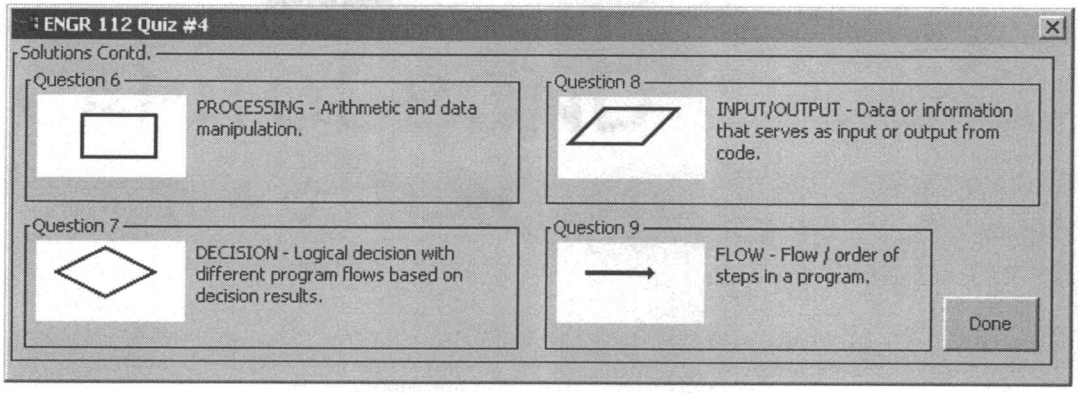


Figure 11: Quiz solutions

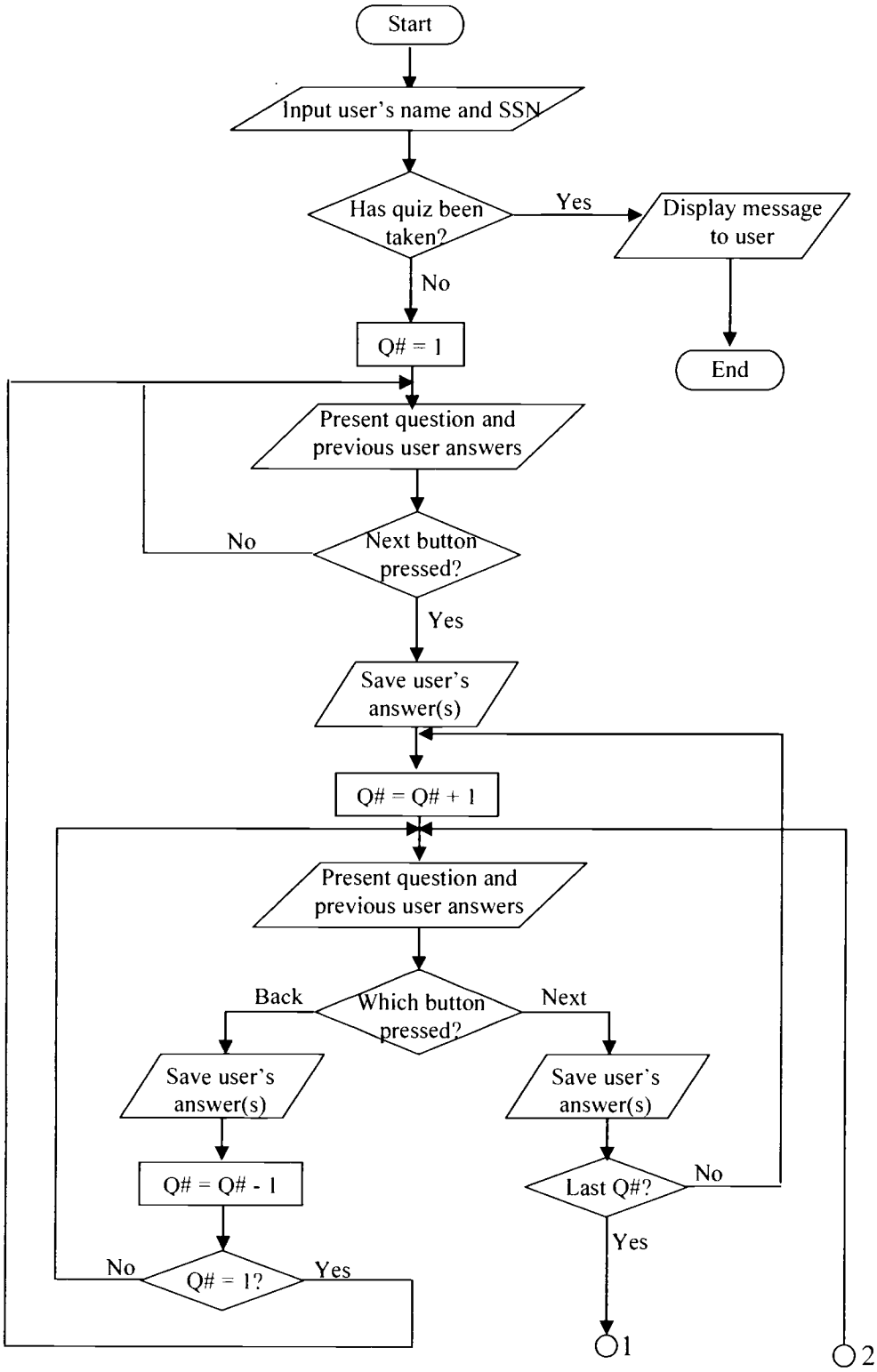


Figure 12: Quiz application flow chart

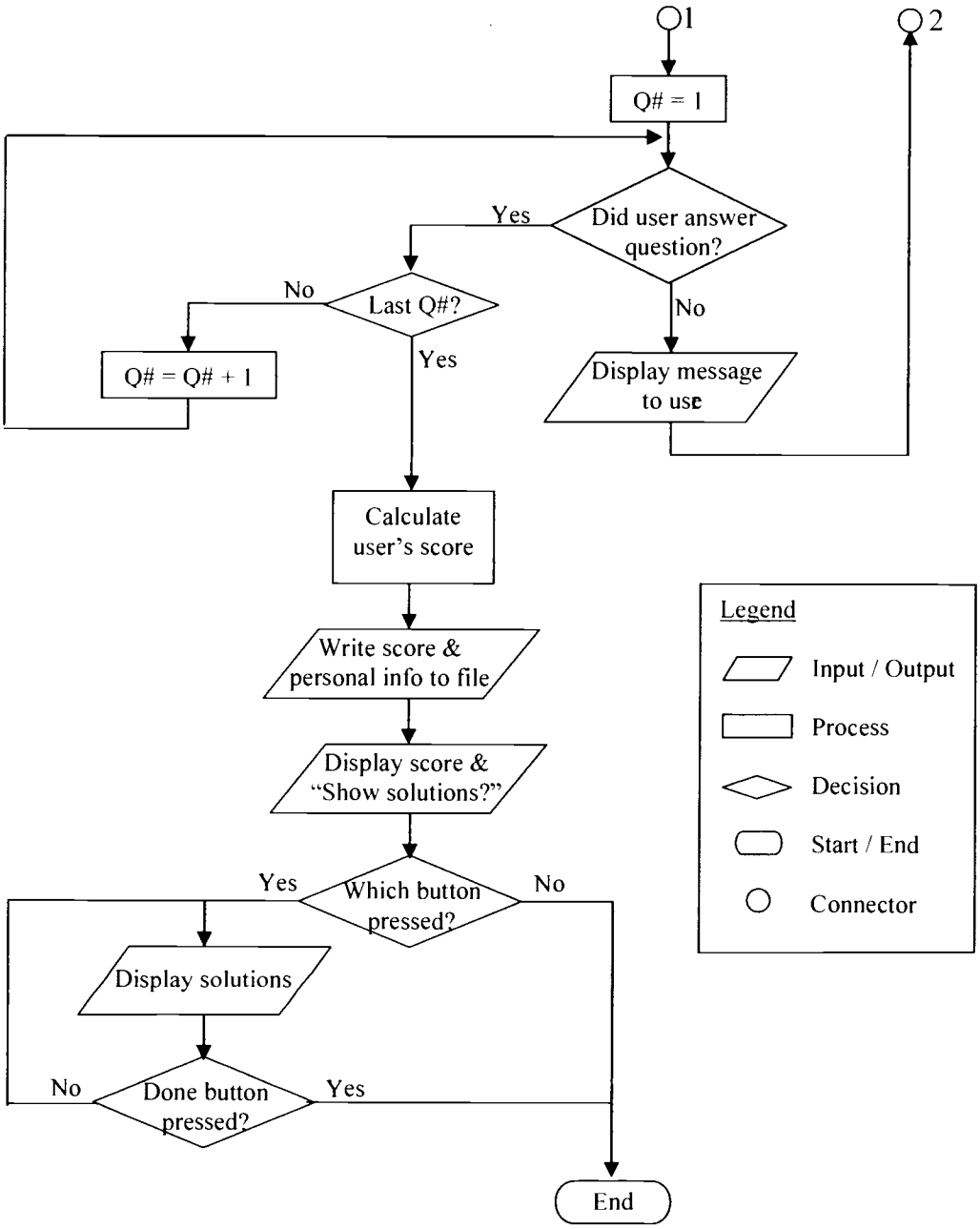


Figure 12 (continued)

After the quiz was completed, students logged on to a secure university website that provided support for file submittal. They selected the assignment for which they wanted to submit the file (see Figure 13) and then browsed the handheld PC to find the .csv file that contained their personal information, score, and answers to the quiz questions (Figure 14). This file was then submitted and saved in a folder labeled with the student's user name on a university server to which only the course instructor and teaching assistants had access.

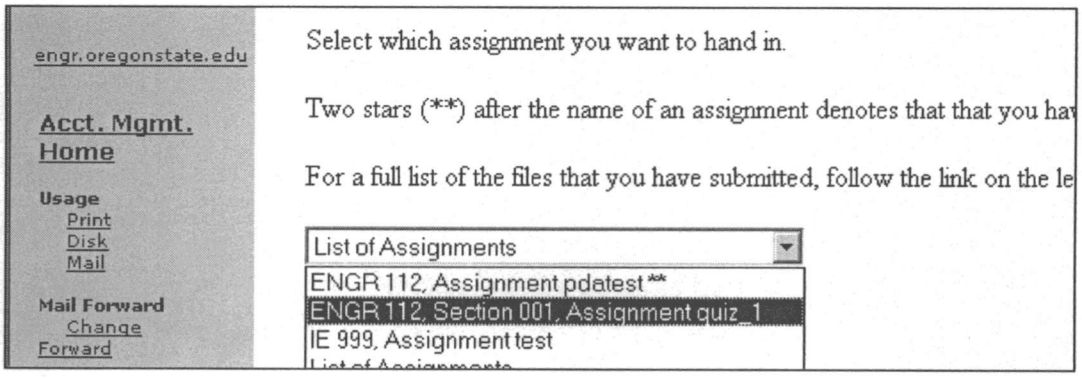


Figure 13: Submitting a file on the web

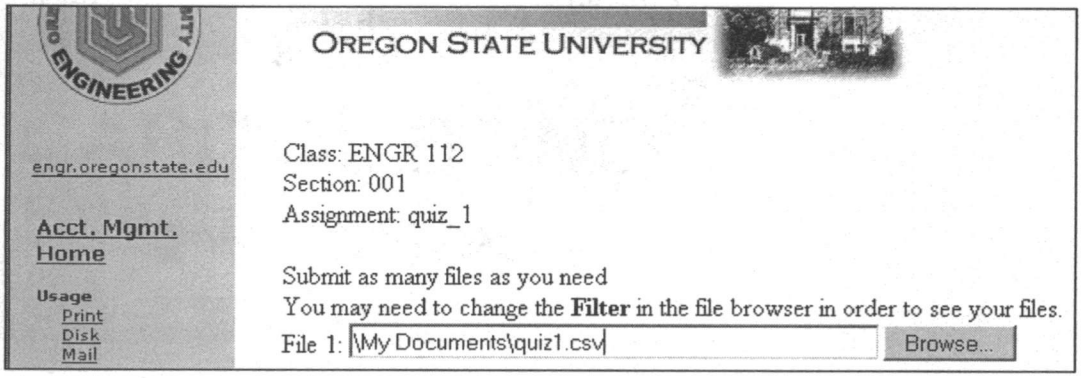


Figure 14: Selection of file to submit

4.5. Analysis Methods

4.5.1. Path analysis

Path analysis (a special case of structural equation modeling) is a method for providing direct and indirect estimates of the magnitude and significance of hypothesized causal relationships between sets of variables (Webley and Lea, 1997). Path diagrams are used to graphically display the different variables and the proposed direction of causality between them. Path coefficients, which are standardized multiple regression coefficients (beta weights), are calculated to quantify the strength of the relations between each pair of variables.

It is important to note that path analysis is not intended to deduce causal relations: it is useful in testing theory rather than in generating it. In addition, some assumptions underlie the application of path analysis: relationships are linear, additive, and causal, and variables are measurable on an interval scale (Land, 1969). Another assumption is that residuals are not correlated among themselves or with the system variables; this implies that all relevant variables are included in the system (Kerlinger and Pedhazur, 1973).

Path analysis has become a basic research tool in the social sciences. It is useful for handling multiple real-life variables, and for this reason it was used in this study. Other statistical tools were ruled out on the grounds that they were less appropriate for this type of research. Analysis of variance (ANOVA), for example, is utilized to compare the means of three or more treatment levels, where here only two groups are always compared. It is also often applied to engineering and scientific problems and was therefore deemed unfit for this study.

4.5.2. Model

Table 9 provides a summary of the independent and the dependent variables that were measured in this study. It also lists how each variable was evaluated.

Table 9: Independent and dependent variables

Metric	Evaluation Method
Independent variables	
• Quiz administration method	• Two administration methods: PDA-based and paper-and-pencil
• Computer anxiety	• Five computer anxiety survey items
Dependent variables	
• Effectiveness	• Quiz score • Five mental workload survey items
• Efficiency	• Time to complete quiz
• Satisfaction	• Five overall satisfaction survey items; seven additional system characteristics survey items following the PDA-based quiz

T-tests were used to compare means where data was assumed to be normally distributed. Paired sets of survey data, which prior research has shown to violate the normality assumption (see section 4.3.4), were compared using the Wilcoxon signed rank test, the non-parametric complement to the paired t-test. Linear regression was used to assess relationships between multiple variables. Linear regression estimations are based on three assumptions: normally distributed residuals, common variance, and independent errors. The existence of these conditions was established before analyzing the results. The models are presented in Table 10.

Table 10: Models

	Hypothesis	Method	Model
1a.	Effectiveness, measured as quiz scores (score) and mental workload (MW), is higher for PDA based quizzes than for paper-and-pencil quizzes.	Paired t-test Wilcoxon signed rank test	Score: $\mu_{PDA} > \mu_{paper}$ MW: $median_{PDA} < median_{paper}$
1b.	PDA based quizzes are more efficient than paper-and-pencil quizzes as measured by quiz completion time (time).	Paired t-test	Time: $\mu_{PDA} < \mu_{paper}$
1c.	User satisfaction (USAT) is higher for the PDA-based quiz than for the paper-and-pencil quiz.	Wilcoxon signed rank test	USAT: $median_{PDA} > median_{paper}$

Table 10 (continued)

	Hypothesis	Method	Model
2a.	Effectiveness (score, MW) is negatively correlated with computer anxiety (CA) for PDA-based quizzes. No relationship exists between computer anxiety and effectiveness in paper-and-pencil quizzes (QAM: quiz administration method).	Linear regression	$\mu\{\text{score} \mid \text{QAM}, \text{CA}\} = \beta_0 + \beta_1 \cdot \text{QAM} + \beta_2 \cdot \text{CA} + + \beta_3(\text{QAM} \times \text{CA})$ $\mu\{\text{MW} \mid \text{QAM}, \text{CA}\} = \beta_0 + \beta_1 \cdot \text{QAM} + \beta_2 \cdot \text{CA} + + \beta_3(\text{QAM} \times \text{CA})$
2b.	User satisfaction (USAT) is negatively correlated with computer anxiety (CA) for PDA-based quizzes. No relationship exists between computer anxiety and satisfaction in paper-and-pencil quizzes (QAM: quiz administration method).	Linear regression	$\mu\{\text{USAT} \mid \text{QAM}, \text{CA}\} = \beta_0 + \beta_1 \cdot \text{QAM} + \beta_2 \cdot \text{CA} + + \beta_3(\text{QAM} \times \text{CA})$
3.	User satisfaction (USAT) is positively correlated with effectiveness (score, MW) for both PDA based and paper-and-pencil quizzes (QAM: quiz administration method).	Linear regression	$\mu\{\text{USAT} \mid \text{QAM}, \text{score}\} = \beta_0 + \beta_1 \cdot \text{QAM} + \beta_2 \cdot \text{score} + + \beta_3(\text{QAM} \times \text{score})$ $\mu\{\text{USAT} \mid \text{QAM}, \text{MW}\} = \beta_0 + \beta_1 \cdot \text{QAM} + \beta_2 \cdot \text{MW} + + \beta_3(\text{QAM} \times \text{MW})$

4.6. Experimental Design

The one-group pretest-posttest design was used in this study. This is a pre-experimental design in which one group is subjected to a treatment and observed before and after this treatment (Campbell and Stanley, 1963). In this study, the treatment is the administration of a PDA-based quiz. Students taking an academic course were observed twice: after taking a paper-and-pencil quiz, which may be considered a control treatment, and after taking a PDA-based quiz. The observation took the form of a survey, which the students filled out following the completion of the quiz. Students' quiz scores and the time it took them to complete the quiz were also recorded.

4.6.1. Participants

The participants for this study were College of Engineering students or students considering engineering as a possible degree option at Oregon State University. These students were taught engineering problem solving using computers, how to use spreadsheet tools (Microsoft Excel), and basic programming skills (Microsoft Visual Basic) in Engineering Orientation II, an engineering introductory course for freshmen. This course was identical to the one from which participants were recruited to the pilot study in which the surveys' reliabilities were evaluated. The full study was completed in the academic term following the pilot study. In a class of 38 students, 92% were under 23 years of age, 87% were males, and 79% were Caucasian. Prior to taking the survey, an informed consent form was distributed and read to the participants; this form stated that participation in the experiment was voluntary. Students were not compensated in any way for their participation.

4.6.2. Instrument

The surveys used to evaluate PDA-based and paper-and-pencil quizzes were similar to those used in the pilot study (see section 4.3.2.). In the surveys following the PDA-based quiz, the questions were modified to refer to PDAs rather than computers for scales evaluating user satisfaction. The number of items evaluating overall user satisfaction was reduced from seven to five; these five items were identical to those used in the survey following the paper-and-pencil quiz. This change was made in order to make the two surveys comparable. The five computer anxiety items used in the survey following the PDA-based quiz were added to the paper-and-pencil quiz survey, again, to make the two surveys comparable (see Appendix C). In addition to the surveys, quiz scores and the time needed to complete the quizzes were recorded. In order to compare students' quiz scores, the last four digits of their social security number were also noted.

4.6.3. Procedure

During the second week of classes, each student received a kit that contained an HP Jornada 720 along with a power cord, wireless LAN card, and a docking cradle used to synchronize files with a desktop computer. The students were allowed to keep the PDAs for the duration of the term. PDAs were used in class to solve engineering problems and take quizzes. In addition, students could follow the instructor's lectures by downloading lecture slides to the PDAs.

During the ten-week term, five quizzes were administered at the beginning of the lecture session. Of these, students took two PDA-based quizzes (in weeks 2 and 8) and three paper-and-pencil quizzes (in weeks 4, 7, and 9). Paper-and-pencil quizzes were similar to the PDA-based quizzes in their format: multiple-choice questions, in which students selected one (or more) answer. After the completion of their second paper-and-pencil quiz (in week 7) and second PDA-based quiz (in week 8), students were asked to fill out the corresponding survey. They were also

asked to record the time at which they began taking the paper-and-pencil quiz, and the time at which the quiz was finished (these times were electronically recorded for the PDA-based quiz). The paper-and-pencil quiz after which surveys were administered tested students' knowledge of economic analysis (see Appendix D). The material covered in the PDA-based quiz after which surveys were administered was basic programming terminology (see Appendix D).

5. Results

Table 11 summarizes the descriptive statistics of the paper-and-pencil quiz, the PDA-based quiz, and the surveys that followed them. A total of 34 students took the paper-and-pencil quiz and 29 completed the corresponding survey; 30 students took the PDA-based quiz and 26 filled out its survey.

The students reacted favorably to both quiz administration methods. They rated both quizzes as demanding low mental workload (less than 35 on a scale of 0 to 100). Average satisfaction scores over 3.0 indicated that students were fairly satisfied with both quiz types. In addition, they were satisfied with the PDA based quiz's system characteristics (such as reliability and speed). Their computer anxiety was relatively low (over 4 out of 5, with high ratings indicating a low level of anxiety), regardless of quiz type.

Table 11: Survey results: Mean scores and standard deviations

		Paper-and-Pencil Quiz		**PDA-based Quiz**	
	Scale	**Mean**	**STD**	**Mean**	**STD**
Effectiveness – Quiz score	0-15	13.59	1.40	13.23	1.98
Effectiveness – Mental workload	0- 100	29.62	14.30	33.84	17.06
Efficiency – Time to complete quiz		4:31	1:42	3:55	1:25
Satisfaction	1-5	3.80	0.58	3.69	0.80
Satisfaction with system characteristics	1-5			3.99	0.65
Computer anxiety	1-5	4.26	0.81	4.25	0.77

5.1. Reliability Evaluation

To ensure the surveys' reliability, Cronbach's alpha was calculated again. The data from the survey following the paper-and-pencil quiz and the survey following the PDA-based quiz was aggregated when survey items were identical. Cronbach's alpha for each item group is presented in Table 12. All alphas were higher than 0.7, the minimum recommended by Nunnally (1978).

Table 12: Cronbach's alpha for survey items

	Number of Items	**n**	**Cronbach's Alpha**
User satisfaction	5	52	0.74
Satisfaction with system characteristics	7	22	0.87
Mental workload	5	54	0.79
Computer anxiety	5	55	0.90

5.2. Hypothesis Checking

5.2.1. Hypothesis 1

Hypothesis 1 states that effectiveness, measured as the students' quiz scores and mental workload, efficiency, measured by quiz completion time, and user satisfaction, are higher for the PDA-based quiz than for the paper-and-pencil quiz.

Quiz scores and mental workload ratings were not significantly different for the two quizzes, therefore effectiveness was not affected by the quiz administration method. Students completed the paper-and-pencil quiz in 4 minutes and 41 seconds, on average, and the PDA-based quiz in 3 minutes and 54 seconds. The time spent on the quizzes was significantly lower for the PDA-based quiz, therefore it was more efficient than the paper-and-pencil quiz. Satisfaction ratings for the two quiz types were not significantly different. The results for the paired data are presented in Table 13.

Table 13: Hypothesis 1 results

	n	**Paper-and-Pencil Quiz Mean**	**PDA-Based Quiz Mean**	**Statistic**
Effectiveness – Quiz score	29	13.72	13.24	t = 1.260
Effectiveness – Mental workload	21	27.90	35.43	Z = -1.737
Efficiency – Time to complete quiz	29	4:41	3:54	t = 2.353*
Satisfaction	22	3.85	3.60	Z = -0.809

* One-sided p-value < 0.05

5.2.2. Hypothesis 2

According to the second hypothesis, both effectiveness and user satisfaction with PDA-based quizzes are negatively correlated with computer anxiety: as computer anxiety increases, effectiveness and satisfaction decrease. No relationship exists between computer anxiety and effectiveness in paper-and-pencil quizzes. Likewise, no relationship exists between computer anxiety and user satisfaction in paper-and-pencil quizzes.

5.2.2.1. Effect of Quiz Administration Method and Computer Anxiety on Effectiveness Measured as Quiz Scores

There was no evidence that the quiz administration method-computer anxiety interaction significantly affected quiz scores (two-sided p-value = 0.122, t-test). There was no evidence that the quiz administration method was associated with score, after accounting for computer anxiety (two-sided p-value = 0.483, t-test). There was convincing evidence that computer anxiety was associated with the quiz score (two-sided p-value < 0.001, t-test). A one-unit increase in the computer anxiety scale was associated with an estimated 0.433 points decrease in the mean quiz score (95% confidence interval from –0.594 to –0.272). Since high computer anxiety ratings indicate a low level of anxiety (e.g. an item rating of 5 would be indicative of lower computer anxiety than a rating of 4), the relationship between computer anxiety and quiz scores is positive: the more anxiety the student experienced, the higher was his or her quiz score, regardless of whether the quiz was paper-and-pencil or PDA-based. This data is displayed in Figure 15 (only part of the 0-15 quiz score scale is displayed, since all scores were above 8 points).

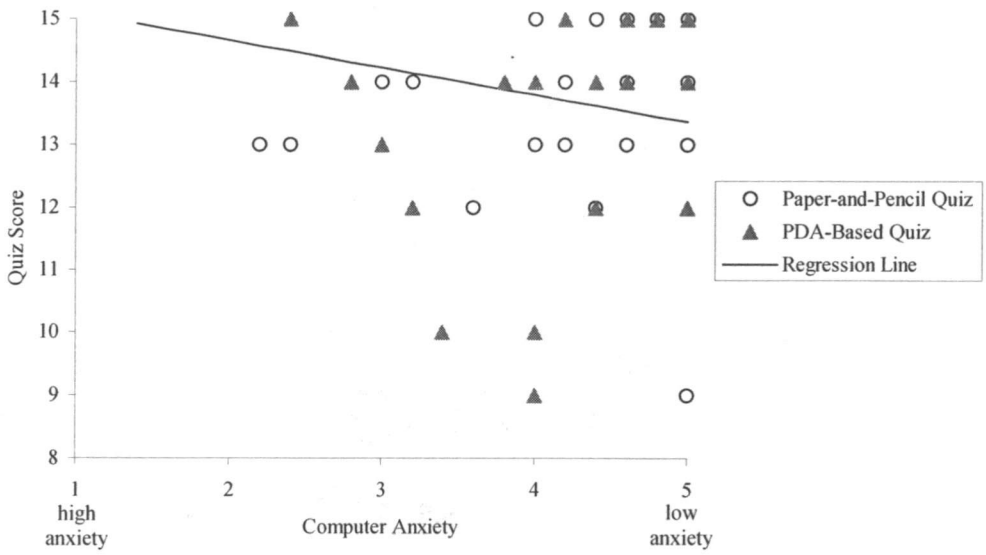


Figure 15: Scatter plot of quiz score as a function of quiz administration method and computer anxiety

5.2.2.2. Effect of Quiz Administration Method and Computer Anxiety on Effectiveness Measured as Mental Workload

There was no evidence that the quiz administration method-computer anxiety interaction significantly affected mental workload ratings (two-sided p-value = 0.583, t-test). There was no evidence that the quiz administration method was associated with mental workload, after accounting for computer anxiety (two-sided p-value = 0.393, t-test). There was evidence that computer anxiety was associated with mental workload (two-sided p-value = 0.028, t-test). A one-unit increase in the computer anxiety scale was associated with an estimated 1.743 units decrease in the mental workload scale (95% confidence interval from –3.292 to –0.194). Since high computer anxiety ratings indicate a low level of anxiety, the relationship between computer anxiety and mental workload is positive: the more anxiety the student experienced, the more workload he or she felt, regardless of whether the quiz was paper-and-pencil or PDA-based. In other words, effectiveness, associated with a low mental workload, is negatively correlated with

computer anxiety. This data is displayed in Figure 16 (only part of the 0-100 mental workload scale is displayed, since all ratings were below 70).

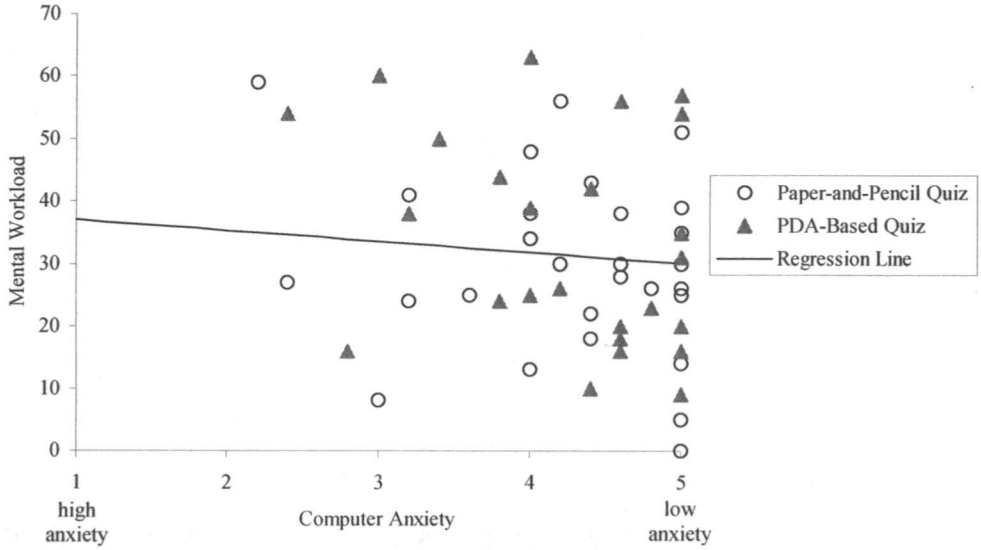


Figure 16: Scatter plot of mental workload as a function of quiz administration method and computer anxiety

5.2.2.3. Effect of Quiz Administration Method and Computer Anxiety on User Satisfaction

There was no evidence that the quiz administration method-computer anxiety interaction significantly affected user satisfaction (two-sided p-value = 0.102, t-test). There was no evidence that the quiz administration method was associated with user satisfaction, after accounting for computer anxiety (two-sided p-value = 0.544, t-test). There was no evidence that computer anxiety was associated with user satisfaction (two-sided p-value = 0.092, t-test). This data is displayed in Figure 17.

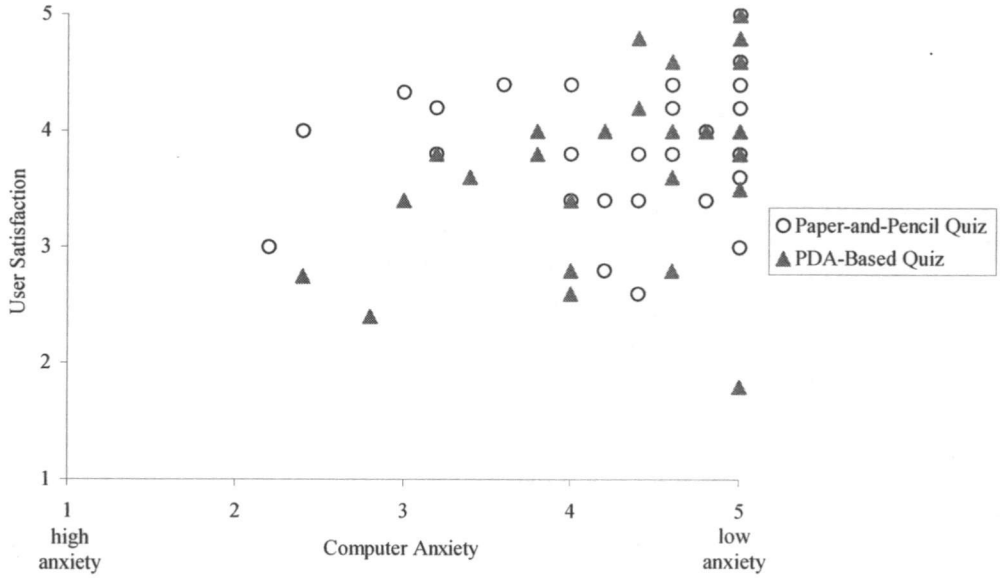


Figure 17: Scatter plot of user satisfaction as a function of quiz administration method and computer anxiety

For all three linear regression models, normal probability plots of the residuals were examined to ensure that the normality assumption holds. In addition, residual plots were examined to ensure that the residual variance homogeneity assumption holds.

5.2.3. Hypothesis 3

The third hypothesis states that user satisfaction is positively correlated with effectiveness for both PDA-based and paper-and-pencil quizzes.

5.2.3.1. Effect of Quiz Administration Method and Effectiveness Measured as Quiz Scores on User Satisfaction

There was no evidence that the quiz administration method-quiz score (effectiveness) interaction significantly affected satisfaction (two-sided p-value = 0.884, t-test). There was no evidence that the quiz administration method was associated with user satisfaction, after accounting for quiz score (two-sided p-

value = 0.596, t-test). There was no evidence that user satisfaction was associated with quiz score (two-sided p-value = 0.158, t-test). This data is displayed in Figure 18 (only part of the 0-15 quiz score scale is displayed, since all scores were above 8 points).

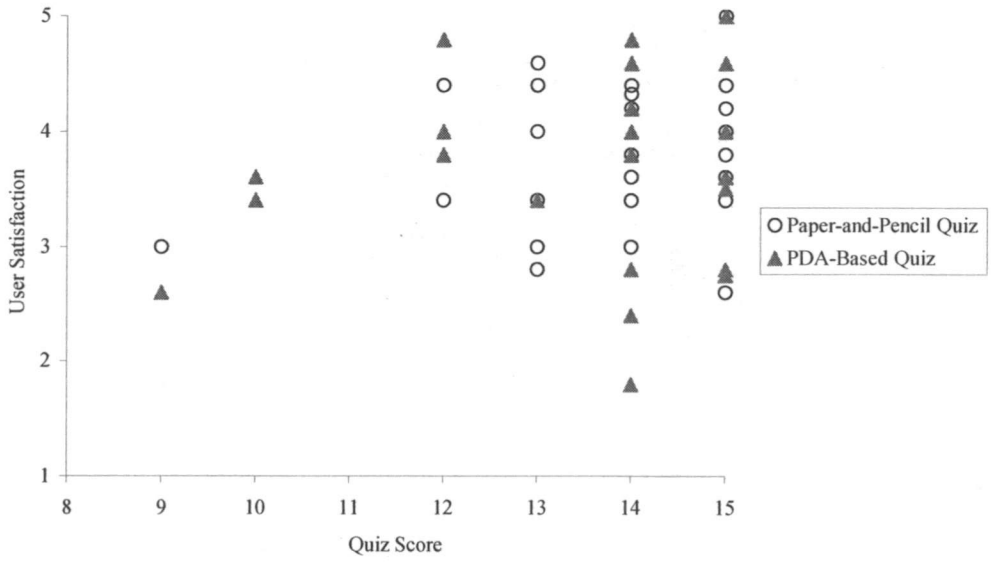


Figure 18: User satisfaction as a function of quiz administration method and quiz score

5.2.3.2. Effect of Quiz Administration Method and Effectiveness Measured as Mental Workload on User Satisfaction

There was no evidence that the quiz administration method-mental workload (effectiveness) interaction significantly affected satisfaction (two-sided p-value = 0.923, t-test). There was no evidence that the quiz administration method was associated with user satisfaction, after accounting for mental workload (two-sided p-value = 0.891, t-test). There was convincing evidence that user satisfaction was associated with mental workload (two-sided p-value < 0.001, t-test). A one-unit increase in the mental workload scale was associated with an estimated 0.021 units decrease in the user satisfaction scale (95% confidence interval from –0.032 to –

0.010). Higher mental workload was associated with lower satisfaction, regardless of whether the quiz was paper-and-pencil or PDA-based. Thus, user satisfaction is positively correlated with effectiveness measured as mental workload. This data is displayed in Figure 19 (only part of the 0-100 mental workload scale is displayed, since all ratings were below 70).

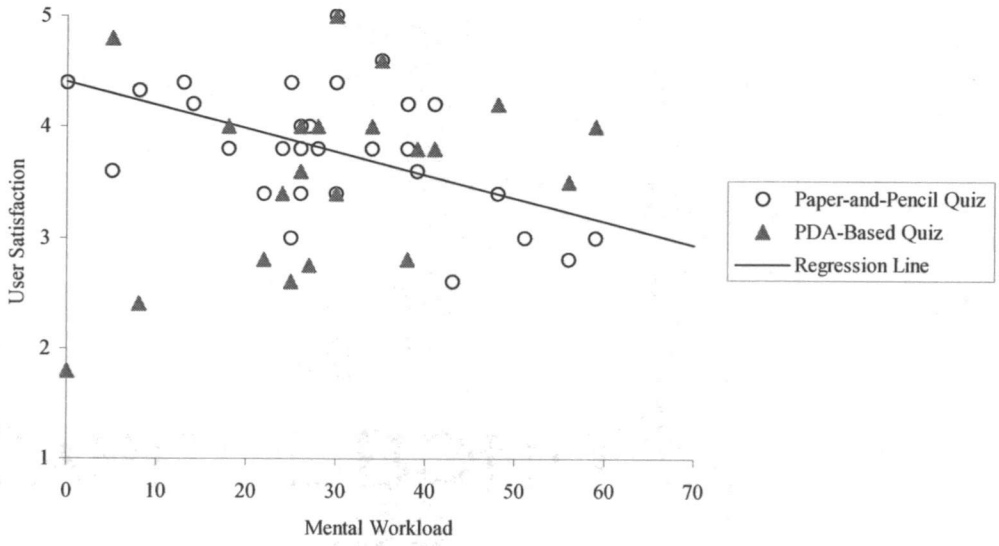


Figure 19: User satisfaction as a function of quiz administration method and mental workload

For both linear regression models, normal probability plots of the residuals were examined to ensure that the normality assumption holds. In addition, residual plots were examined to ensure that the residual variance homogeneity assumption holds.

5.2.4. Summary

The thesis hypotheses were partially confirmed. No relationships were found between some of the independent and dependent variables. There were opposing findings in the second hypothesis. Table 14 summarizes the findings.

Table 14: Hypotheses and findings

	Hypothesis	Findings
1a.	Effectiveness, measured as quiz scores and mental workload, is higher for PDA-based quizzes than for paper-and-pencil quizzes.	Quiz scores and mental workload ratings were not significantly different for the two quizzes, therefore effectiveness was not affected by the quiz administration method.
1b.	PDA based quizzes are more efficient than paper-and-pencil quizzes as measured by quiz completion time.	The time spent on the quizzes was significantly lower for the PDA-based quiz, therefore it was more efficient than the paper-and-pencil quiz.
1c.	User satisfaction is higher for the PDA based quiz than for the paper-and-pencil quiz.	Satisfaction ratings for the two quiz types were not significantly different.
2a.	Effectiveness (quiz score and mental workload) is negatively correlated with computer anxiety for PDA-based quizzes. No relationship exists between computer anxiety and effectiveness in paper-and-pencil quizzes.	As computer anxiety increased, effectiveness increased (quiz scores increased), regardless of quiz type. As computer anxiety increased, effectiveness decreased (mental workload increased), regardless of quiz type.

Table 14 (continued)

	Hypothesis	Findings
2b.	User satisfaction is negatively correlated with computer anxiety for PDA-based quizzes. No relationship exists between computer anxiety and satisfaction in paper-and-pencil.	No relationship existed between computer anxiety and satisfaction for either paper-and-pencil or PDA-based quizzes.
3.	User satisfaction is positively correlated with effectiveness (quiz score and mental workload) for both PDA-based and paper-and-pencil quizzes.	No relationship existed between quiz score and satisfaction for neither quiz type. User satisfaction was positively correlated with effectiveness measured as mental workload, regardless of quiz type.

5.3. Path Diagrams

The path diagrams are presented in Figures 20 and 21. In Figure 20 effectiveness is measured using quiz scores, and in Figure 21 it is measured using the mental workload scale. Bold arrows signify significant relationships. One-sided p-values are noted for significant relationships. For continuous variables, a (+) indicates positive relationships (an increase in the independent variable is associated with an increase in the dependent variable) and (-) indicates negative relationships.

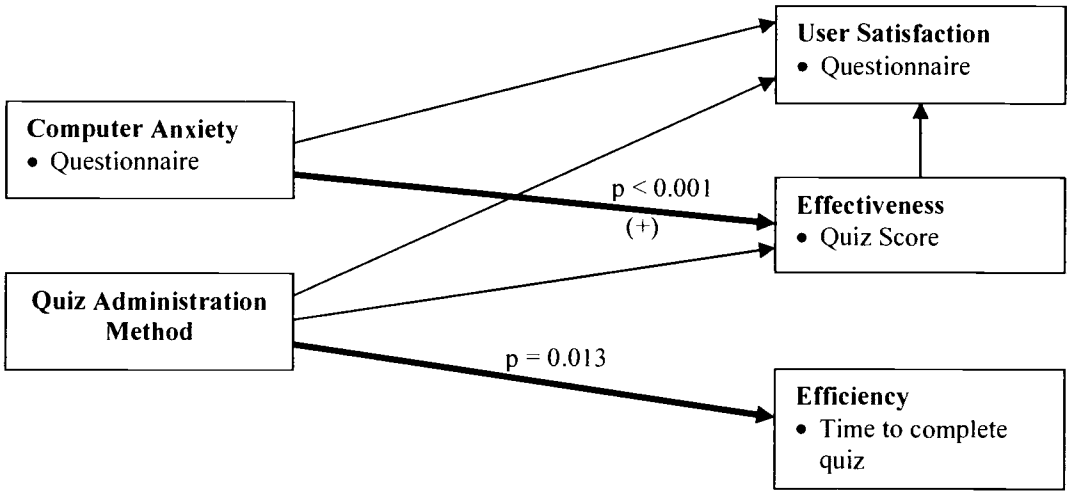


Figure 20: Path diagram for effectiveness measured as quiz scores

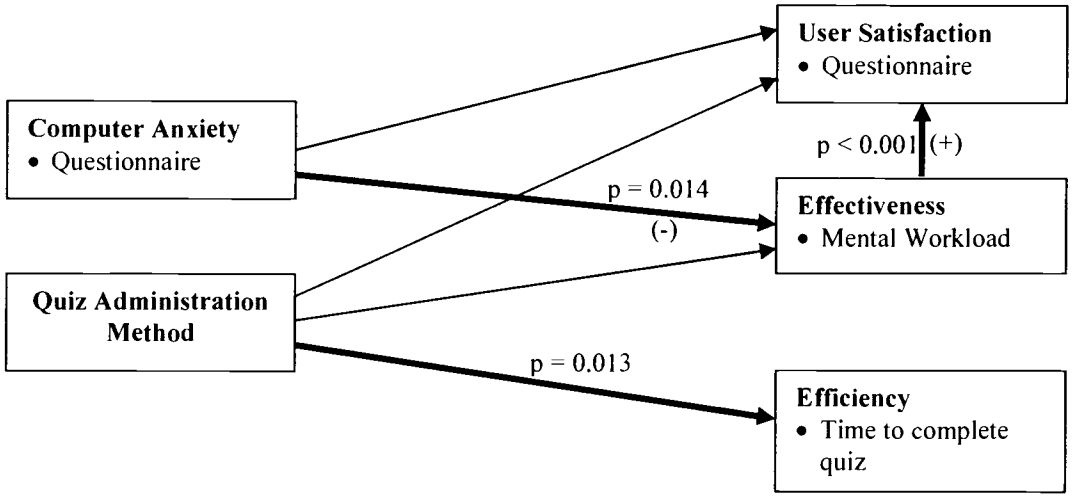


Figure 21: Path diagram for effectiveness measured as mental workload.

5.4. Demographic Comparisons

The dependent variables were compared in terms of gender, age, and ethnicity, to determine whether different populations reacted differently to the PDA-based quiz. Gender comparisons are summarized in Table 15. There were no significant differences between males and females, with the exception of the paper-and-pencil quiz satisfaction ratings. Females were more satisfied with this quiz than males. However, the female sample size was small (4 to 5 participants). Additional research with a larger sample size may be required to substantiate these findings.

Table 15: Means and statistics by gender

		Males		**Females**		
	Quiz Type	**Mean**	**n**	**Mean**	**n**	**Statistic**
Effectiveness – Quiz score	Paper-and-pencil	13.63	27	14	5	t = -0.555
	PDA-based	13.55	22	12.60	5	t = 0.626
Effectiveness – Mental workload	Paper-and-pencil	28.88	24	33.20	5	Z = -0.838
	PDA-based	31.30	20	42.50	4	Z = -1.202
Efficiency – Time to complete quiz	Paper-and-pencil	4:20	27	5:24	5	t = -1.271
	PDA-based	3:59	22	3:20	5	t = 1.273
Satisfaction	Paper-and-pencil	3.67	24	4.40	5	Z = -2.503*
	PDA-based	3.69	21	3.75	4	Z = -0.224
Computer anxiety	Paper-and-pencil	4.23	24	4.44	5	Z = -0.382
	PDA-based	4.22	21	4.65	4	Z = -1.021

* One-sided p-value < 0.05

Table 16 summarizes age comparisons. Students were divided into two age categories: younger and older than 23 years of age. No significant differences were found between the two age groups. However, only three participants were older than 23. Additional research with a larger sample size may be required to find significant differences, should they exist.

Table 16: Means and statistics by age

		Younger than 23		**Older than 23**		
	Quiz Type	**Mean**	**n**	**Mean**	**n**	**Statistic**
Effectiveness – Quiz score	Paper-and-pencil	13.66	29	14.00	3	t = -0.414
	PDA-based	13.25	24	14.33	3	t = -0.888
Effectiveness – Mental workload	Paper-and-pencil	30.19	26	24.67	3	Z = -0.430
	PDA-based	33.77	21	34.33	3	Z = -0.481
Efficiency – Time to complete quiz	Paper-and-pencil	4:26	29	5:00	3	t = -0.516
	PDA-based	3:57	24	3:10	3	t = 0.880
Satisfaction	Paper-and-pencil	3.78	26	3.93	3	Z = -0.469
	PDA-based	3.74	22	3.33	3	Z = -0.126
Computer anxiety	Paper-and-pencil	4.22	26	4.60	3	Z = -0.583
	PDA-based	4.21	22	4.53	3	Z = -0.341

Table 17 presents data for white and non-white students. Non-white students (Asian, Middle-Eastern, and Hispanic or Latino Americans) were grouped

together. Average PDA-based quiz scores were 8.3% lower for white students; no other significant differences were found between the two groups. Only 6-8 students identified themselves as belonging to ethnic categories other than white, thus further research with a larger sample size may be required to find significant differences, should they exist.

Table 17: Means and statistics by ethnicity

		White		**Non-White**		
	Quiz Type	**Mean**	**n**	**Mean**	**n**	**Statistic**
Effectiveness – Quiz score	Paper-and-pencil	13.54	24	14.13	8	t = 1.057
	PDA-based	13.05	20	14.29	7	t = 2.181*
Effectiveness – Mental workload	Paper-and-pencil	30.00	22	28.43	7	Z = -0.281
	PDA-based	33.28	18	32.83	6	Z = -0.267
Efficiency – Time to complete quiz	Paper-and-pencil	4:42	24	3:52	8	t = -1.180
	PDA-based	4:02	20	3:21	7	t = -1.072
Satisfaction	Paper-and-pencil	3.71	22	4.08	7	Z = -1.362
	PDA-based	3.70	19	3.70	6	Z = 0.00
Computer anxiety	Paper-and-pencil	4.26	22	4.26	7	Z = -0.026
	PDA-based	4.35	19	4.10	6	Z = -0.519

* One-sided p-value < 0.05

5.5. Validity and Reliability

The one-group pretest-posttest design is widely used in educational research (Campbell and Stanley, 1963), but several variables jeopardize its internal and external validity. For an experiment to have internal validity, it needs to be shown that the treatments, and not extraneous factors, were responsible for changes in dependent variables. There are several factors that present alternative, plausible explanations to these changes:

- History. Many external change-producing events may have occurred between the two quizzes. One such change is the fact that the quizzes themselves were different: they covered different material, the number of questions was different, and they were given at different points in time.
- Maturation. Biological and psychological processes such as tiredness, hunger, boredom, etc. may have influenced the study participants. This factor most likely did not have a major effect on study results, since the two quizzes were taken on the same day of the week at the same time. Therefore, biological and psychological processes should not have been very different on the two survey days.
- Testing. Student answers to the survey items may have differed from one survey to the next due to the effect of replying to similar items twice.

Two factors strengthen the internal validity of the study:

- Selection. The same students took the quizzes and answered the surveys. Had two groups each taken one quiz and completed its respective survey, a claim could be made that some unmeasured factor made one group different from the other and accounted for the measured changes.

- Mortality. Paired statistical tests compared each student's responses to the first and second surveys. If two groups had each taken one quiz and completed its respective survey at different points in time, a loss of respondents could have occurred. As the term progresses, some students drop out of courses; usually, these are the weaker students. Thus, differences between survey results could be attributed to the first group being stronger than the second group.

External validity establishes the domain to which study findings can be generalized: can the observed effect be generalized to other populations, settings, and treatment and measurement variables? Though often internal and external validity are not independent, i.e. increasing one decreases the other, the aim is to use an experimental design that is strong in both types of validity. Three factors limit this study's external validity:

- Interaction of testing and the PDA-based quiz. A pretest (in this study, the first survey) often changes participants' attitudes as manifested in the posttest (the second survey). Consequently, the results cannot be generalized to the universal student population that has not been exposed to the sensitizing effect of the pretest.

  All together, the students who participated in this study were asked to complete five surveys throughout the term in the engineering introduction course alone. This may have caused them to fill out the surveys in an offhand manner, without giving proper attention to each survey item. As a result, the responses may not be representative of the general student population.
- Interaction of selection and the PDA-based quiz. The effects demonstrated in this study may hold only for the population from which the participants were selected. For example, middle school students may react differently to PDA-based quizzes than university students.

- Reactive arrangements. The artificiality of the experimental setting and students' knowledge that they are participating in an experiment may affect the results. This effect is similar to the effect of the testing and the PDA-based quiz interaction, and can be avoided by keeping the students from knowing that they are taking part in an experiment. The students received the PDAs after the first week of classes and kept them for the duration of the term. They were used in class exercises and two quizzes were administered on the PDAs. Therefore the treatment and the control, which are the PDA-based quiz and the paper-and-pencil quiz, respectively, were usual classroom events taking place at plausible times. The surveys, however, were not disguised (e.g. embedded into regular quizzes or exams), so the effect of reactive arrangements cannot be entirely ruled out.

Factors affecting the validity and reliability of the dependent variables are summarized in Table 18. A minus (-) indicates the factor was not addressed in the experimental design, a plus (+) indicates that the factor is controlled, and a zero (0) indicates that the factor is not applicable to the experimental design.

Table 18: Validity and reliability of dependent variables

Variable	Internal Validity		External Validity		Reliability	
Effectiveness – Quiz score	-	History	0	Testing-PDA-based quiz interaction		
	+	Maturation				
	0	Testing	-	Selection-PDA-based quiz interaction		
	+	Selection				
	+	Mortality	0	Reactive arrangements		
Effectiveness – Mental workload	-	History	-	Testing-PDA-based quiz interaction	+	Cronbach's alpha = 0.79
	+	Maturation				
	-	Testing	-	Selection-PDA-based quiz interaction		
	+	Selection				
	+	Mortality	-	Reactive arrangements		
Efficiency – Time to complete quiz	-	History	0	Testing-PDA-based quiz interaction		
	+	Maturation				
	0	Testing	-	Selection-PDA-based quiz interaction		
	+	Selection				
	+	Mortality	0	Reactive arrangements		
Satisfaction	-	History	-	Testing-PDA-based quiz interaction	+	Cronbach's alpha = 0.74
	+	Maturation				
	-	Testing	-	Selection-PDA-based quiz interaction		
	+	Selection				
	+	Mortality	-	Reactive arrangements		

6. Discussion

6.1. Conclusions

Several conclusions can be drawn from this study's results:

The paper-and-pencil quiz and the PDA-based quiz were equally effective. This is in accordance with the findings of Perkins (1995), who compared computer-based and paper-and-pencil tests, although other studies have documented different findings (Chin and Donn, 1991). The paper-and-pencil quiz and the PDA-based quiz were also equally satisfactory to the students.

The PDA-based quiz was more efficient than the paper-and-pencil quiz, since it was completed in less time. However, this result should be carefully interpreted given that the two quizzes differed in structure and material covered. In addition, errors may have been introduced into the paper-and-pencil quiz time calculations, since start and stop times were recorded by the students (and not electronically, as they were in the PDA-based quiz) and were rounded to the nearest minute.

Students exhibited relatively low computer anxiety. This is not surprising: it is unlikely that they would have chosen to study engineering (as most of them have), a profession that entails much contact with computers, had they found them threatening. This finding held true across gender, age, and ethnicity. Results showing ethnicity to have no effect on computer anxiety support previous research (Bowers and Bowers, 1996, and Gilroy and Desai, 1986). Likewise, similar to the findings of this study, most studies examining student populations characterized by a narrow age range found no relationship between age and computer anxiety (Gilroy and Desai, 1986, and Maurer, 1994). Bowers and Bowers (1996) did find a

positive correlation between the two factors, but they surveyed social science students, whilethe participants in this study were mostly engineering students. Regarding gender, research on its relationship to computer anxiety has obtained mixed results (Worthington and Zhao, 1999), thus the results of this study confirmed some studies and refuted others.

Two conflicting trends were observed in the effectiveness-computer anxiety relationship. When effectiveness was measured as quiz scores, this relationship was positive: students who experienced higher anxiety were more effective (received higher scores), regardless of quiz type. This is in contradiction with Perkins' (1995) finding that lower anxiety predicts higher test scores, and with studies showing no relationship between computer anxiety and test scores. On the other hand, when effectiveness was measured as mental workload, the relationship was negative: students who experienced more anxiety were less effective (felt a higher workload), regardless of quiz type. These findings would indicate that perhaps effectiveness is not a one-dimensional construct. For example, it is possible that high mental workload is indicative of greater, rather than lower, effectiveness for certain tasks. A quiz should not be too easy, therefore students should have to put an effort into it, in order to enable the instructor to use quiz scores to rank students' performance (Kehoe, 1995).

Two conflicting trends were observed in the satisfaction-effectiveness relationship. When effectiveness was measured as quiz scores, no relationship was found between the two variables. When effectiveness was measured as mental workload, this relationship was positive: students who were more effective (experienced a lower workload) were more satisfied with the quiz, regardless of quiz type. This may again be explained by the multidimensionality of the effectiveness construct, if mental workload and quiz scores measure its different aspects.

In evaluating the usability of new software, its effectiveness may be measured in several different ways, e.g. the number of errors made by users when performing a task, the number of tasks users complete in a certain amount of time, the subjective workload the users feel when using the software, etc. In this study, effectiveness was quantified by measuring the number of errors (quiz scores) and the mental workload, and findings were different for each metric. One may infer, then, that each metric measures a different aspect of effectiveness; thus, for each software to be evaluated, the researcher needs to select and measure those metrics that are most relevant to its uses. This conclusion is also true for the evaluation of efficiency, satisfaction, and any other usability metric. For an application measuring students' performance, it is possible that greater emphasis should be placed on quiz scores (and less on mental workload ratings), as this is the most direct measurement of its effectiveness.

Causality cannot be inferred for the computer anxiety-effectiveness and effectiveness-satisfaction relationships. It is possible, for example, that students who effectively complete a quiz (e.g. receive a high score) will also be more satisfied with it. It is also possible that students who are satisfied with a quiz will, as a result, complete it more effectively. A third possibility is that a third factor, for example motivation, drives both effectiveness and satisfaction.

Females were more satisfied with the paper-and-pencil quiz than males, and PDA-based quiz scores were slightly lower for white students. No other differences were found between different demographic groups.

6.2. Implications

A usability comparison of paper-and-pencil and PDA-based quizzes has found the latter to be equal, if not superior, to the former. The effort students put into taking the quiz was the same, regardless of administration method, and scores were not affected. In addition, different demographic groups performed almost equally well in both quiz types (white students' PDA-based quiz scores were slightly lower than those of the other ethnic groups). Computer anxiety was not affected by the quiz type. For these reasons, as well as other advantages to both students (e.g. real-time scoring) and teachers (e.g. spending less time on grading), PDAs are an attractive test administration option for schools and universities. The use of handheld devices in education has drawbacks as well, which must be considered when deciding whether to invest in them. First, a test application suitable for the PDA needs to be developed. Issues such as what type of questions will be included in the test need to be decided upon. For example, open-ended questions are not recommended, since text input is often limited in handheld devices (Mohageg, 1999). In addition, issues such as security, how to prevent students from cheating, how to make the exam available to students, and how to make students' answers and scores available to the instructor must be resolved. Finally, if the instructor needs to devote more time to administrative matters related to processing PDA-based exams (e.g. connectivity problems, software issues, etc.) than those experienced when administering and grading a paper-and-pencil exam, then PDAs may not be a feasible alternative. In other words, a robust, comprehensive solution needs to be designed, taking into account educational, administrative, system, financial, and usability requirements.

6.3. Future Work

One of the limitations of this research is its experimental design: the same student population took both quizzes, therefore changes in the dependent variables may be explained by external factors (see section 5.5.). This design jeopardizes the internal validity of this study. To achieve a more valid experimental design, two randomly assigned student groups should take the same quiz at the same time, with one group taking a paper-and-pencil version of the quiz and the other taking a PDA-based version. One way to achieve this design is to randomly select half of the students in a class to take a paper-and-pencil quiz while the other half takes the same quiz on a PDA.

The reactions of males and females, different age groups, and different ethnicities to the quizzes were compared in this study. However, some groups were underrepresented – females, non-white ethnic populations, and older students. Furthermore, the sample population was homogeneous in that it only included university engineering students. To generalize this study's findings to broader populations, e.g. high school students or non-technical university students, it would need to be replicated with a larger, more diverse student sample.

In the pilot study, comparisons were made between computer-based and paper-and-pencil quizzes in terms of satisfaction and effectiveness (mental workload). Students rated the computer-based quiz as more satisfactory and less effort demanding than the paper-and-pencil quiz. Schools and universities that need to choose between setting up a computer lab and investing in handheld devices and wireless technology would benefit from an experiment comparing computer-based, PDA-based, and paper-and-pencil quizzes.

Bibliography

1. Abramovici, M., and Klußmann, N. (1994). Graphical user interface style guide for mobile communication services. In Second International Conference on Intelligence in Broadband Services and Networks (pp. 89-97). Berlin, Germany: Springer-Verlag.

2. Battiste, V., and Bortolussi, M. (1988). Transport pilot workload: a comparison of two subjective techniques. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 150-154). Santa Monica, CA: Human Factors and Ergonomics Society.

3. Bellamy, R., Swart, C., Kellogg, W.A., Richards, J., and Brezin, J. (2001). Designing an e-grocery application for a palm computer: usability and interface issues. IEEE Personal Communications, 8(4), 60-64.

4. Besterfield-Sacre, M., Moreno, M., Shuman, L.J., and Atman, C.J. (1999, June). Comparing entering freshman engineers: institutional differences in student attitudes. Presented at the American Society for Engineering Education Annual Conference, Charlotte, NC.

5. Bjork, S., Holmquist, L.E., Redstrom, J., Bretan, I., Danielsson, R., Karlgren, J., and Franzen, K. (1999). WEST: a web browser for small terminals. In Proceedings of 12th Annual Symposium on User Interface Software and Technology (pp. 187-196). New York, NY: ACM.

6. Booth, J.F. (1998). The user interface in computer-based selection and assessment: Applied and theoretical problematics of an evolving technology. International Journal of Selection & Assessment, 6(2), 61-81.

7. Borland Developer Network (1999). Delphi 3 file types with descriptions. From http://community.borland.com/article/0,1410,16552,00.html.

8. Bowers, D.A., and Bowers, V.M. (1996). Assessing and coping with computer anxiety in the social science classroom. Social Science Computer Review, 14(4), 439-443.

9. Branaghan, R.J. (2001). Human factors issues in the design of handheld wireless devices. In Proceedings of the 2001 International Conference on High-Density Interconnect and Systems Packaging (pp. 37-41). Washington, DC: International Microelectronics and Packaging Society.

10. Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. Personal and Ubiquitous Computing, 6(3), 188-205.

11. Campbell, D.T., and Stanley, J.C. (1963). Experimental and quasi-experimental designs for research. Boston, MA: Houghton Mifflin Company.

12. Chen, F., Myers, B., and Yaron, D. (2000). Using Handheld Devices for Tests in Classes. (Tech. Report CMU-CS-00-152). Pittsburgh: Carnegie Mellon University, School of Computer Science.

13. Chin, C.H.L., and Donn, J.S. (1991). Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. Educational & Psychological Measurement, 51(3), 735-745.

14. Chin, J.P., Diehl, V.A., and Norman, K.L. (1988). Development of a tool measuring user satisfaction of the human-computer interface. In Proceedings of ACM CHI'88 Conference on Human Factors in Computing Systems (pp. 213-218). New York, NY: Association for Computing Machinery.

15. Cook, R.P. (2000). The national classroom project – an experience report. In Proceedings of the 30th ASEE/ISEE Frontiers in Education Conference (pp. T1E/1-6). Champaign, IL: Stripes Publishing.

16. Dean, K. (2002). Study: PDAs good for education. From http://www.wired.com/news/school/0,1383,56297,00.html.

17. DeVellis, R.F. (1991). Scale development: theory and applications. Newbury Park, CA: Sage Publications.

18. Devlin, S.J., and Dong, H.K. (1993). Selecting a scale for measuring quality. Marketing Research, 5(3), 12-17.

19. Dimock, P.H. (1991). The effects of format differences and computer experience on a computer-administered test. Measurement & Evaluation in Counseling & Development, 24(3), 119-126.

20. Europemedia (2002). PDA market predicted to have steady sales growth for next 4 years. From http://www.europemedia.net/shownews.asp?ArticleID=13163.

21. Gilroy, F.D., and Desai, H.B. (1986). Computer anxiety: sex, race, and age. International Journal of Man-Machine Studies, 25, 711-719.

22. Gressard, C., and Loyd, B. (1986). Validation studies of a new computer attitude scale. Association for Educational Data Systems Journal, 18(4), 295-301.

23. Harper, B., Slaughter, L., and Norman, K. (1997). Questionnaire administration via the WWW: A validation & reliability study for a user satisfaction questionnaire. Presented at the Proceedings of WebNet 97: International Conference on the WWW, Internet, and Intranet, Toronto, ON, Canada.

24. Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad, A.L., and Christ, R.E. (1992). Comparison of four subjective workload rating scales. Human Factors, 34(4), 429-439.

25. Hudgins, B. (2001). Leveraging handheld technology in the classroom. Technological Horizons in Education, 29(5), 46-47.

26. Huff, K.L., and Sireci, S.G. (2001). Validity issues in computer-based testing. Educational Measurement: Issues and Practice, 20(3), 16-25.

27. Igbaria, M., Schiffman, S.J., and Wieckowski, T.J. (1994). The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. Behaviour & Information Technology, 13(6), 349-361.

28. Institute of Electrical and Electronics Engineers (1990). IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY: Author.

29. Kabara, J., Krishnamurthy, P., and Weiss, M. (2000). Use of wireless computers in the undergraduate and graduate classroom. In 30th ASEE/IEEE Frontiers in Education Conference (pg. F2D-1). New York, NY: Institute of Electrical and Electronics Engineers.

30. Kamba, T., Elson, S.A., Harpold, T., Stamper, T., and Sukaviriya, P. (1996). Using small screen space more efficiently. In Proceedings of the 1996 Conference on Human Factors in Computing Systems (pp. 383-390). New York, NY: ACM.

31. Kawamoto, D. (2003). PDAs need more work to thrive. From http://zdnet.com.com/2100-1103-986172.html.

32. Kehoe, J. (1995). Basic item analysis for multiple-choice tests. Practical Assessment, Research & Evaluation, 4(10). From http://edresearch.org/pare/getvn.asp?v=4&n=10.

33. Kerlinger, F.N., and Pedhazur, E.J. (1973). Multiple Regression in Behavioral Research. New York, NY: Holt, Rinehart and Winston, Inc.

34. Kline, P. (1994). An easy guide to factor analysis. London: Routledge.

35. LaLomia, M.J., and Sidowski, J.B. (1990). Measurements of computer satisfaction, literacy, and aptitudes: a review. International Journal of Human-Computer Interaction, 2(3), 231-253.

36. Land, K.C. (1969). Principles of Path Analysis. In E.F. Borgatta (Ed.), Sociological Methodology (pp. 3-37). San Francisco, CA: Josey Bass, Inc.

37. Loyd, B.H., and Gressard, C. (1984). Reliability and Factorial Validity of Computer Attitude Scales. Educational and Psychological Measurement, 44(2), 501-505.

38. Masui, N., Okazaki, T., and Tonomura, Y. (1999). Proposal and evaluation of user interface for personal digital assistants. In Proceedings of 8th International Conference on Human Computer Interaction (pp. 968-972). Mahwah, NJ: Lawrence Erlbaum Associates.

39. Maurer, M.M. (1994). Computer anxiety correlates and what they tell us: a literature review. Computers in Human Behavior, 10(3), 369-376.

40. Mayhew, D. (1992). Principles and guidelines in software user interface design. Englewood Cliffs, NJ: Prentice Hall.

41. McGookin, D.K., and Brewster, S.A. (2001). FISHEARS – The design of a multimodal focus and context system. In 15th Annual Conference of the Human-Computer Interaction Group of the British Computer Society (pp. 1-4). Toulouse, France: Cepadues-Editions.

42. Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23, 12-23.

43. Mohageg, M.F. (1999). User interface design in the post-PC era. In Proceedings of 8th International Conference on Human Computer Interaction (pp. 1137-1142). Mahwah, NJ: Lawrence Erlbaum Associates.

44. Nash, J.B., and Moroz, P.A. (1997). An examination of the factor structures of the computer attitude scale. Journal of Educational Computer Research, 17(4), 341-356.

45. Nielsen, J. (1993). Usability engineering. Boston, MA: Academic Press.

46. North Atlantic Treaty Organization, Research and Technology Organization. (2001). NATO guidelines on human engineering testing and evaluation (Tech. Report RTO-TR-021). Neuilly-Sur-Seine Cedex, France: Author.

47. Nunnally, J.C. (1978). Psychometric theory. New York, NY: McGraw Hill.

48. Nyberg, M., Bjork, S., Goldstein, M., and Redström, J. (2001). Handheld applications design: merging information appliances without affecting usability. In Proceeding of 8th International Conference on Human-Computer Interactions (pp. 391-398). Amsterdam, Netherlands: IOS Press.

49. PDA cortex (2003). Perception of PDA functionality key to growth in PDA market. From http://www.rnpalm.com/growth_pda_market.htm.

50. Perkins, R.F. (1995). Using hypermedia programs to administer tests: Effects on anxiety and performance. Journal of Research on Computing in Education, 28(2), 209-220.

51. Raub, A.C. (1981). Correlates of computer anxiety in college students. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

52. Reid, G. B., and Nygren, T. E. (1988). The Subjective Workload Assessment Technique: a scaling procedure for measuring mental workload. In P.A. Hancock and N. Meshkati (Eds.), Human Mental Workload (pp. 185-218). Amsterdam, The Netherlands: Elsevier Science Publishers.

53. Rekimoto, J. (1998). A multiple device approach for supporting whiteboard-based interactions. In Proceedings of Conference on Human Factors in Computing Systems (pp. 344-351). New York, NY: ACM.

54. Robertson, S., Wharton, C., Ashworth, C., and Franzke, M. (1996). Dual device user interface design: PDAs and interactive television. In Proceedings of CHI 96: Human Factors in Computing Systems (pp. 79-86). New York, NY: ACM.

55. Root, R.W., and Draper, S. (1983). Questionnaires as a software evaluation tool. In Proceedings of CHI 83: Conference on Human Factors in Computing Systems (pp. 83-87). Boston, MA: ACM.

56. Sanders, M.S., and McCormick, E.J. (1993). Human Factors in Engineering and Design. New York, NY: McGraw-Hill.

57. Shneiderman, B. (1998). Designing the user interface: strategies for effective human-computer interaction. Reading, MA: Addison-Wesley.

58. Shotsberger, P. and Vetter, R. (2001). Teaching and learning in the wireless classroom, Computer, 34(3), 110-111.

59. Smith, W.J. (1996). ISO and ANSI ergonomic standards for computer products: a guide to implementation and compliance. Upper Saddle River, NJ: Prentice-Hall.

60. SPSS Inc. (1991). SPSS base 10.0 applications guide. Chicago, IL: SPSS, Inc.

61. Taivalsaari, A. (1999). The event horizon user interface model for small devices (Technical Report TR-99-74). Palo Alto, CA: Sun Microsystems Laboratories.

62. Webley, P., and Lea, S. (1997). Path Analysis. From http://www.ex.ac.uk/~SEGLea/multvar2/pathanal.html.

63. Whiteside, J., Bennett, J., and Holtzblatt, K. (1988). Usability engineering: our experience and evolution. In M.G. Helander, T. K. Landauer, and P. Prabhu (Eds.), Handbook of human computer interaction (pp. 791-818). New York: North Holland.

64. Worthington, V.L., and Zhao, Y. (1999). Existential computer anxiety and changes in computer technology: what past research on computer anxiety has missed, Journal of Educational Computer Research, 20(4), 299-315.

65. Xerox: Palo Alto Research Lab (2002). Projects: Information Interfaces. From http://www2.parc.com/istl/projects/uir/projects/ii.html.

Appendices

Appendix A: Questionnaires

Questionnaire for User Interface Satisfaction

Overall reactions to the software:

1. Terrible/Wonderful
2. Frustrating/Satisfying
3. Dull/Stimulating
4. Difficult/Easy
5. Inadequate power/ Adequate power
6. Rigid/Flexible

Screen:

7. Characters on the computer screen – Hard to read/Easy to read
8. Highlighting on the screen simplifies task – Not at all/Very much
9. Organization of information on screen – Confusing/Very clear
10. Sequence of screens – Confusing/Very clear

Terminology and system information:

11. Use of terms throughout system – Inconsistent/Consistent
12. Computer terminology is related to the task you are doing – Never/Always
13. Position of messages on screen – Inconsistent/Consistent
14. Messages on screen which prompt user for input – Confusing/Clear
15. Computer keeps you informed about what it is doing – Never/Always
16. Error messages – Unhelpful/Helpful

Learning:

17. Learning to operate the system – Difficult/Easy
18. Exploring new features by trial and error – Difficult/Easy
19. Remembering names and use of commands – Difficult/Easy
20. Tasks can be performed in a straightforward manner – Never/Always
21. Help messages on the screen – Unhelpful/Helpful
22. Supplemental reference materials – Confusing/Clear

System capabilities:

23. System speed – Too slow/Fast enough
24. System reliability – Unreliable/Reliable
25. System tends to be – Noisy/Quiet
26. Correcting your mistakes– Difficult/Easy
27. Experienced and inexperienced users' needs are taken into consideration – Never/Always

NASA Task Load Index

Measure	Scale Endpoints	Description
Mental demand	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical demand	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Computer Anxiety Scale

1. I would feel comfortable working with a computer
2. Working with a computer would make me very nervous
3. It wouldn't bother me at all to take computer courses
4. I feel aggressive and hostile toward computers
5. I do not feel threatened when others talk about computers
6. I would feel at ease in a computer class
7. Computers make me feel uneasy and confused
8. Computers make me feel uncomfortable
9. Computers do not scare me at all
10. I get a sinking feeling when I think of trying to use a computer

Appendix B: Pilot Study Surveys

Pilot Study Survey for Computer-Based Quiz

Computer Quiz Survey

Please select the ONE response that best describes your opinion with respect to taking the quiz on a computer. If you are unsure about an item, leave it blank.

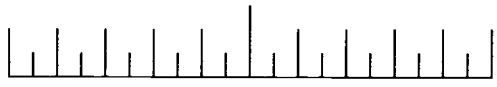
1.	Was the experience of taking the quiz on a computer terrible or wonderful?	Terrible	1	2	3	4	5	Wonderful
2.	Was it difficult or easy to take the quiz on a computer?	Difficult	1	2	3	4	5	Easy
3.	Was the experience of taking the quiz on a computer frustrating or satisfying?	Frustrating	1	2	3	4	5	Satisfying
4.	Was the experience of taking the quiz on a computer dull or stimulating?	Dull	1	2	3	4	5	Stimulating
5.	Was the computer-based quiz rigid or flexible?	Rigid	1	2	3	4	5	Flexible
6.	How difficult or easy was it to read the characters on the computer screen?	Hard to read	1	2	3	4	5	Easy to read
7.	Was the organization of the screen design confusing or clear?	Confusing	1	2	3	4	5	Very clear
8.	Was the screens' sequence confusing or clear?	Confusing	1	2	3	4	5	Very clear

9.	Were the messages on the screen which prompted you for input confusing or clear? ..	Confusing	1	2	3	4	5	Clear
10.	Were error messages unhelpful or helpful?	Unhelpful	1	2	3	4	5	Helpful
11.	How difficult or easy was it to learn to operate the computer-based quiz?	Difficult	1	2	3	4	5	Easy
12.	Was the system speed too slow or fast enough?	Too slow	1	2	3	4	5	Fast enough
13.	Was the system unreliable or reliable?	Unreliable	1	2	3	4	5	Reliable
14.	How difficult or easy was it to correct your mistakes?	Difficult	1	2	3	4	5	Easy

Please evaluate the task of taking the quiz on a computer by marking each item at the point that matches your experience.

15. Mental Demand

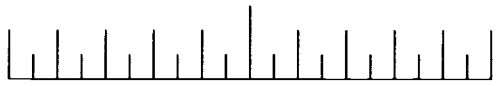
Overall, how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?



Low Mental Demand — High Mental Demand

16. Temporal Demand

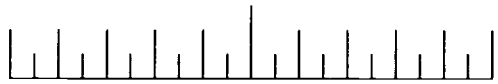
Overall, how much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?



Low Temporal Demand — High Temporal Demand

17. Effort

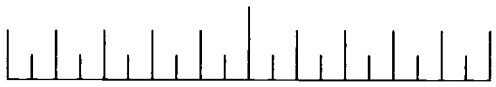
Overall, how hard did you have to work mentally to accomplish your level of performance?



Low Effort — High Effort

18. Frustration Level

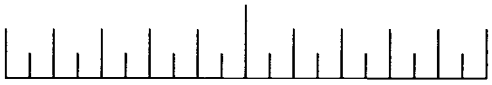
Overall, how insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



Low Frustration Level — High Frustration Level

19. Performance

Overall, how successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?



Good Performance — Poor Performance

Note: Good performance is located on the left-hand side of the scale, poor performance is on the right-hand side.

Please indicate your level of agreement or disagreement for each of the following statements. Circle one number for each statement.

		Strongly Disagree	Some-what Disagree	Neutral	Some-what Agree	Strongly Agree
20.	I feel comfortable working with a computer	1	2	3	4	5
21.	Computers make me feel uneasy and confused	1	2	3	4	5
22.	Working with a computer makes me very nervous	1	2	3	4	5
23.	Computers do not scare me at all	1	2	3	4	5

24. I get a sinking feeling when I think of trying to use a computer	1	2	3	4	5

25. **Gender** (circle one): 1. Male 2. Female

26. **Age** (select one): 1. 23 years old or younger 2. Older than 23 years old

27. **Ethnic identity** (select the one that best applies to you):

a. American Indian or Alaskan Native
b. Black, African American, Non-Hispanic
c. Native Hawaiian or other Pacific Islander
d. White, European American, Non-Hispanic
e. Asian or Asian American
f. Middle Eastern or Middle Eastern American
g. Hispanic or Latino American
h. North African or North African American
i. Other

Thank you very much for your participation!

Pilot Study Survey for Paper-and-Pencil Quiz

Paper Quiz Survey

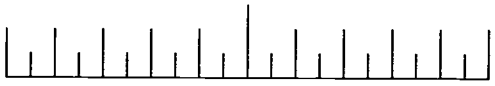
Please select the ONE response that best describes your opinion with respect to taking the quiz on paper. If you are unsure about an item, leave it blank.

1.	Was the experience of taking the quiz on paper terrible or wonderful?	Terrible	1	2	3	4	5	Wonderful
2.	Was it difficult or easy to take the quiz on paper?	Difficult	1	2	3	4	5	Easy
3.	Was the experience of taking the quiz on paper frustrating or satisfying?	Frustrating	1	2	3	4	5	Satisfying
4.	Was the experience of taking the quiz on paper dull or stimulating?	Dull	1	2	3	4	5	Stimulating
5.	Was the paper-based quiz rigid or flexible?	Rigid	1	2	3	4	5	Flexible

Please evaluate the task of taking the quiz on paper by marking each item at the point that matches your experience.

6. Mental Demand

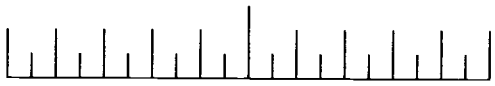
   Overall, how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?



Low Mental Demand — High Mental Demand

7. Temporal Demand

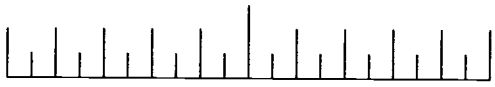
   Overall, how much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?



Low Temporal Demand — High Temporal Demand

8. Effort

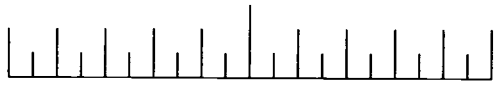
   Overall, how hard did you have to work mentally to accomplish your level of performance?



Low Effort — High Effort

9. Frustration Level

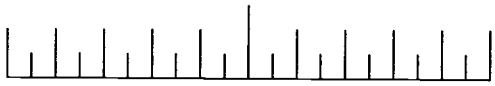
   Overall, how insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



Low Frustration Level — High Frustration Level

10. Performance

    Overall, how successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?



Good Performance — Poor Performance

Note: Good performance is located on the left-hand side of the scale, poor performance is on the right-hand side.

11. **Gender** (circle one): 1. Male 2. Female

12. **Age** (select one): 1. 23 years old or younger 2. Older than 23 years old

13. **Ethnic identity** (select the one that best applies to you):

a. American Indian or Alaskan Native
b. Black, African American, Non-Hispanic
c. Native Hawaiian or other Pacific Islander
d. White, European American, Non-Hispanic
e. Asian or Asian American
f. Middle Eastern or Middle Eastern American
g. Hispanic or Latino American
h. North African or North African American
i. Other

Thank you very much for your participation!

Appendix C: Surveys

Survey for PDA-Based Quiz

PDA Quiz Survey

Last 4 digits of your SSN: _____

Please select the ONE response that best describes your opinion with respect to taking the quiz on a PDA. If you are unsure about an item, leave it blank.

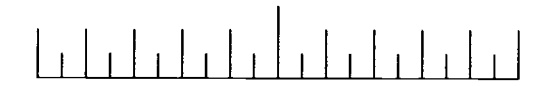
1.	Was the experience of taking the quiz on a PDA terrible or wonderful?	Terrible	1	2	3	4	5	Wonderful
2.	Was it difficult or easy to take the quiz on a PDA? ...	Difficult	1	2	3	4	5	Easy
3.	Was the experience of taking the quiz on a PDA frustrating or satisfying?	Frustrating	1	2	3	4	5	Satisfying
4.	Was the experience of taking the quiz on a PDA dull or stimulating?	Dull	1	2	3	4	5	Stimulating
5.	Was the PDA-based quiz rigid or flexible?	Rigid	1	2	3	4	5	Flexible
6.	How difficult or easy was it to read the characters on the PDA screen?	Hard to read	1	2	3	4	5	Easy to read
7.	Was the organization of the screen design confusing or clear?	Confusing	1	2	3	4	5	Very clear

8.	Was the screens' sequence confusing or clear?	Confusing	1	2	3	4	5	Very clear
9.	Were the messages on the screen which prompted you for input confusing or clear?	Confusing	1	2	3	4	5	Clear
10.	Were error messages unhelpful or helpful?	Unhelpful	1	2	3	4	5	Helpful
11.	Was the system speed too slow or fast enough?	Too slow	1	2	3	4	5	Fast enough
12.	Was the system unreliable or reliable?	Unreliable	1	2	3	4	5	Reliable

Please evaluate the task of taking the quiz on a PDA by marking each item at the point that matches your experience.

13. Mental Demand

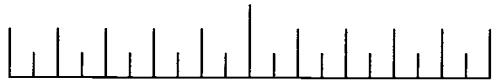
Overall, how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?



Low Mental Demand — High Mental Demand

14. Temporal Demand

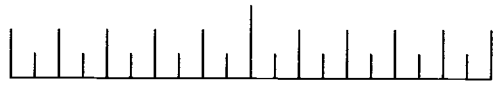
Overall, how much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?



Low Temporal Demand — High Temporal Demand

15. Effort

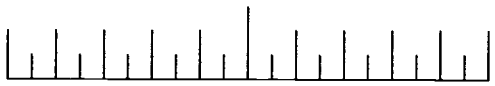
Overall, how hard did you have to work mentally to accomplish your level of performance?



Low Effort — High Effort

16. Frustration Level

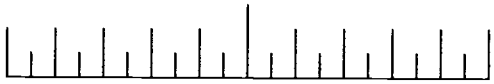
Overall, how insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



Low Frustration Level — High Frustration Level

17. Performance

Overall, how successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?



Good Performance — Poor Performance

Note: Good performance is located on the left-hand side of the scale, poor performance is on the right-hand side.

Please indicate your level of agreement or disagreement for each of the following statements. Circle one number for each statement.

		Strongly Disagree	Some-what Disagree	Neutral	Some-what Agree	Strongly Agree
18.	I feel comfortable working with a computer	1	2	3	4	5
19.	Computers make me feel uneasy and confused ...	1	2	3	4	5
20.	Working with a computer makes me very nervous	1	2	3	4	5
21.	Computers do not scare me at all	1	2	3	4	5

22. I get a sinking feeling when I think of trying to use a computer	1	2	3	4	5

28. **Gender** (circle one): 1. Male 2. Female

29. **Age** (select one): 1. 23 years old or younger 2. Older than 23 years old

30. **Ethnic identity** (select the one that best applies to you):

a. American Indian or Alaskan Native
b. Black, African American, Non-Hispanic
c. Native Hawaiian or other Pacific Islander
d. White, European American, Non-Hispanic
e. Asian or Asian American
f. Middle Eastern or Middle Eastern American
g. Hispanic or Latino American
h. North African or North African American
i. Other

Thank you very much for your participation!

Survey for Paper-and-Pencil Quiz

Paper Quiz Survey

Last 4 digits of your SSN: _____

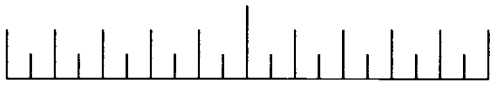
Please select the ONE response that best describes your opinion with respect to taking the quiz on paper. If you are unsure about an item, leave it blank.

1.	Was the experience of taking the quiz on paper terrible or wonderful?	Terrible	1	2	3	4	5	Wonderful
2.	Was it difficult or easy to take the quiz on paper?	Difficult	1	2	3	4	5	Easy
3.	Was the experience of taking the quiz on paper frustrating or satisfying?	Frustrating	1	2	3	4	5	Satisfying
4.	Was the experience of taking the quiz on paper dull or stimulating?	Dull	1	2	3	4	5	Stimulating
5.	Was the paper-based quiz rigid or flexible?	Rigid	1	2	3	4	5	Flexible

Please evaluate the task of taking the quiz on paper by marking each item at the point that matches your experience.

6. Mental Demand

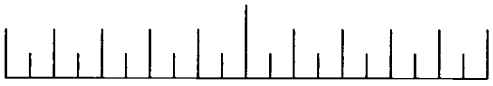
   Overall, how much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?



Low Mental Demand — High Mental Demand

7. Temporal Demand

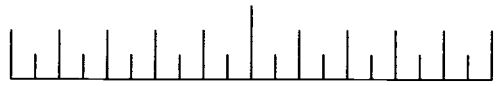
   Overall, how much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?



Low Temporal Demand — High Temporal Demand

8. Effort

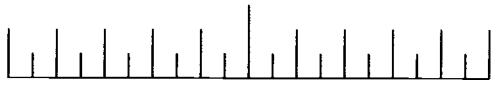
   Overall, how hard did you have to work mentally to accomplish your level of performance?



Low Effort — High Effort

9. Frustration Level

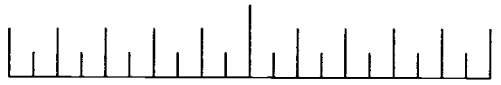
   Overall, how insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?



Low Frustration Level — High Frustration Level

10. Performance

    Overall, how successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?



Good Performance — Poor Performance

Note: Good performance is located on the left-hand side of the scale, poor performance is on the right-hand side.

Please indicate your level of agreement or disagreement for each of the following statements. Circle one number for each statement.

		Strongly Disagree	Some-what Disagree	Neutral	Some-what Agree	Strongly Agree
11.	I feel comfortable working with a computer	1	2	3	4	5
12.	Computers make me feel uneasy and confused ...	1	2	3	4	5
13.	Working with a computer makes me very nervous	1	2	3	4	5
14.	Computers do not scare me at all	1	2	3	4	5
15.	I get a sinking feeling when I think of trying to use a computer	1	2	3	4	5

31. **Gender** (circle one): 1. Male 2. Female

32. **Age** (select one): 1. 23 years old or younger 2. Older than 23 years old

33. **Ethnic identity** (select the one that best applies to you):

a. American Indian or Alaskan Native
b. Black, African American, Non-Hispanic
c. Native Hawaiian or other Pacific Islander
d. White, European American, Non-Hispanic
e. Asian or Asian American
f. Middle Eastern or Middle Eastern American
g. Hispanic or Latino American
h. North African or North African American
i. Other

Thank you very much for your participation!

Appendix D: Quizzes

Paper-and-Pencil Quiz

Quiz #3 May 14, 2003

ENGR 112

Spring 2003

15 points

Name _____

Please record the start and stop time for the quiz.

START TIME:_____

1. A sum of money that is loaned or borrowed is called (1 pt)

   a. Principle b. Interest c. Payment d. Net Present Value

2. A payment that is made for the use of someone else's money is called (1 pt)

   a. Principle b. Interest c. Payment d. Net Present Value

3. Simple interest includes accrued interest? (1 pt)

   a. True b. False

4. Engineering economic analysis is used to evaluate the _____ aspects of a project? (1 pt)

   a. Financial b. Design c. Safety d. Human Resource

5. A dollar today is more valuable than a dollar one year from now. (1 pt)

   a. True b. False

6. The sentence below contains a choice of words in *italics*. Make each of the following statements true by circling the correct words. (4 pts)

   The future value of an investment with a simple interest rate can be calculated by

   x(1+yz) / x*(1+y)^z / x*y*z*

   where x is the *principle / number of years / interest rate* ,

   y is the *principle / number of years / interest rate*,

   and z is the *principle / number of years / interest rate*.

7. With annual compounding, the future value of an investment is the same for simple and compound interest at the end of what **two** years? (2 pts)

   a. Year 0 b. Year 1 c. Year 2 d. Year 3 e. Year n

8. How many times a year is the interest compounded if it is compounded quarterly? (1 pt)

   a. 1 b. 2 c. 4 d. 12

9. How many times a year is the interest compounded if it is compounded annually? (1 pt)

   a. 1 b. 2 c. 4 d. 12

10. How many times a year is the interest compounded if it is compounded semiannually? (1 pt)

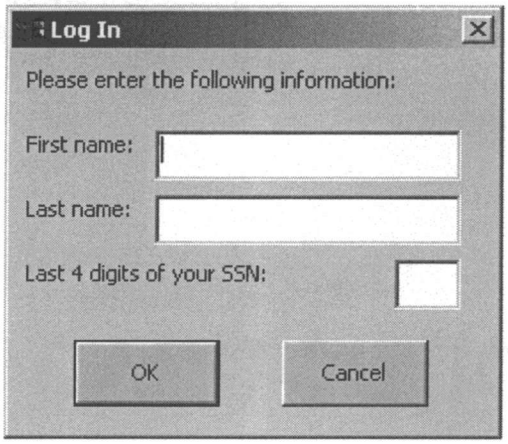
    a. 1 b. 2 c. 4 d. 12

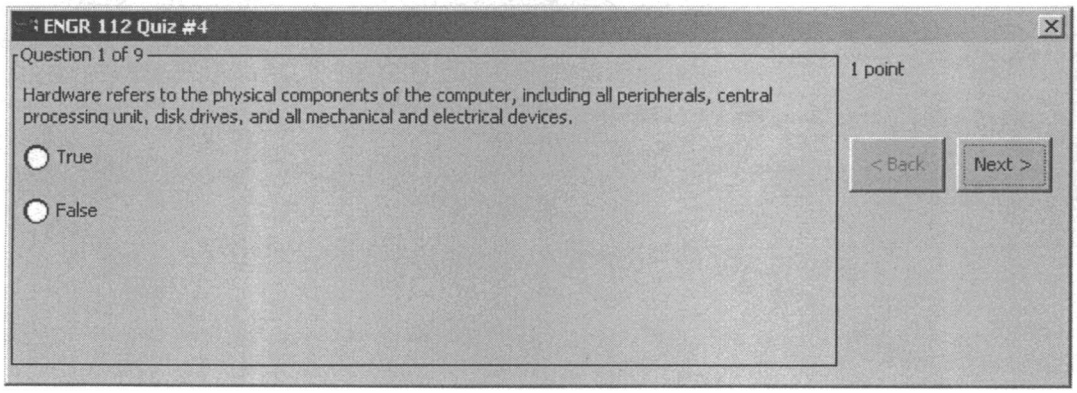
11. How many times a year is the interest compounded if it is compounded monthly? (1 pt)

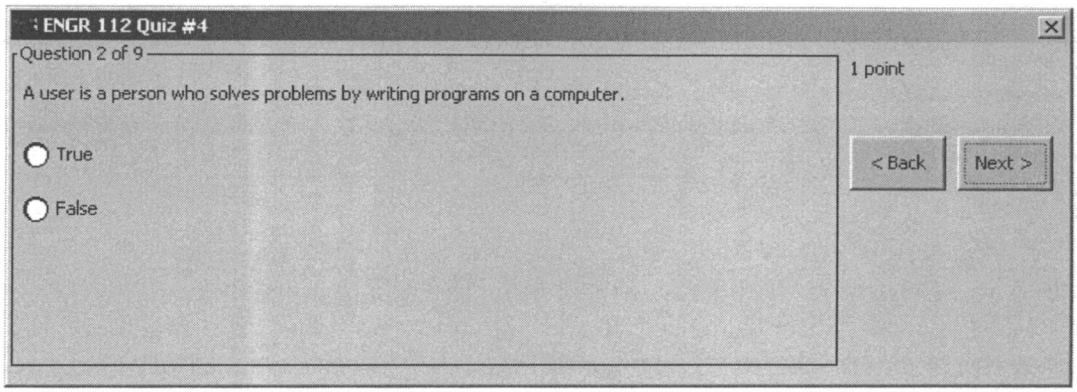
a. 1 b. 2 c. 4 d. 12

STOP TIME:_____

PDA-Based Quiz







ENGR 112 Quiz #4

Question 3 of 9

1 point

What is a GUI?

- General User Interface
- General User Input
- Graphical User Interface
- Graphical User Input

< Back | Next >

ENGR 112 Quiz #4

Question 4 of 9

1 point

A sequence of instructions expressed in a computer language is called:

- GUI
- Class
- Form
- Program

< Back | Next >

ENGR 112 Quiz #4

Question 5 of 9

3 points

Select the THREE functions that nearly all programs have in common:

- [ ] Pseudocode
- [ ] Order
- [ ] Data processing
- [ ] Data input or reading
- [ ] Capability
- [ ] Data output or printing

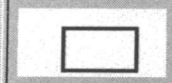
< Back | Next >

ENGR 112 Quiz #4

Question 6 of 9

2 points

Select the name and description for the flow chart symbol using the American National Standard Institute symbology.



Select a name

- Select a name
- Decision
- Flow
- Processing
- Input/Output

- Arithmetic and data manipulation
- Logical decision with different program flows based on decision results
- Data or information that serves as input or output from code
- Flow / order of steps in a program

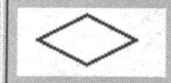
< Back Next >

ENGR 112 Quiz #4

Question 7 of 9

2 points

Select the name and description for the flow chart symbol using the American National Standard Institute symbology.



Select a name

- Select a name
- Decision
- Flow
- Processing
- Input/Output

- Arithmetic and data manipulation
- Logical decision with different program flows based on decision results
- Data or information that serves as input or output from code
- Flow / order of steps in a program

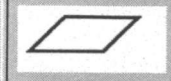
< Back Next >

ENGR 112 Quiz #4

Question 8 of 9

2 points

Select the name and description for the flow chart symbol using the American National Standard Institute symbology.



Select a name

- Select a name
- Decision
- Flow
- Processing
- Input/Output

- Arithmetic and data manipulation
- Logical decision with different program flows based on decision results
- Data or information that serves as input or output from code
- Flow / order of steps in a program

< Back Next >

ENGR 112 Quiz #4

Question 9 of 9

2 points

Select the name and description for the flow chart symbol using the American National Standard Institute symbology.

Select a name

Select a name
Decision
Flow
Processing
Input/Output

- Arithmetic and data manipulation
- Logical decision with different program flows based on decision results
- Data or information that serves as input or output from code
- Flow / order of steps in a program

< Back | Finish

Visual Basic

You received 15 points out of 15. Please submit the file John_Doe.csv located in the My Documents folder. Would you like to see the correct answers?

Yes | No

ENGR 112 Quiz #4

Solutions

Question 1
Hardware refers to the physical components of the computer, including all peripherals, central processing unit, disk drives, and all mechanical and electrical devices - TRUE.

Question 2
A user is a person who solves problems by writing programs on a computer - FALSE.

Question 3
A GUI is a GRAPHICAL USER INTERFACE.

Question 4
A sequence of instructions expressed in a computer language is called a PROGRAM.

Question 5
The three features that nearly all programs have in common: DATA PROCESSING, DATA INPUT OR READING, and DATA OUTPUT OR PRINTING.

Next >

ENGR 112 Quiz #4

Solutions Contd.

Question 6
PROCESSING - Arithmetic and data manipulation.

Question 7
DECISION - Logical decision with different program flows based on decision results.

Question 8
INPUT/OUTPUT - Data or information that serves as input or output from code.

Question 9
FLOW - Flow / order of steps in a program.

Done